

Chapter 4

A Network Analysis of Wikipedia Editors’ Engagement with History: Interests, Identities, Power, and Hierarchy

The English Wikipedia constitutes a huge project that includes a million articles written by millions of editors.¹ It would be impossible to understand and explore how Wikipedians produce historical knowledge without examining Wikipedia on a large scale. In this chapter, I experiment with network analysis and study Wikipedia as a network of interactions between editors and pages. My goal is to investigate the characteristics of Wikipedia users who contribute to pages related to history, the volume of edits they make, their interests, experiences, and education, which prompts them to take part in the Wikipedia community and to produce historical knowledge. As the media scholar José van Dijck has argued, Wikipedia’s success lies in the fact that it can mobilize different types of users to contribute to its contents.² What are the characteristics of those users and how are their characteristics related to their engagement with history? By conducting a network analysis of Wikipedians who edit historical articles, I detect repeating patterns that reveal why Wikipedia editors decide to contribute to pages related to history and what characteristics these editors have.³ If we assume that each Wikipedia page is a network in which several editors participate, is it a balanced and symmetrical network? Or are there a few users who monopolize both the cre-

1 “English Wikipedia,” *Wikipedia*, accessed September 10, 2021, https://en.wikipedia.org/wiki/English_Wikipedia

2 van Dijck, *The Culture of Connectivity*, 136.

3 For the use of networks in the humanities, see Ahnert, et al., *The Network Turn*; David Easley and Jon Kleinberg, ed., *Networks, Crowds, and Markets: Reasoning About a Highly Connected World* (Cambridge University Press, 2010); Roberto Franzosi and John W. Mohr, “New Directions in Formalization and Historical Analysis,” *Theory and Society* 26, nos. 2–3 (1997): 133–60. For some examples of network analysis in the humanities, see Ruth Ahnert and Sebastian Ahnert, “Protestant Letter Networks in the Reign of Mary I: A Quantitative Approach,” *English Literary History* 82, no. 1 (Spring 2015): 1–33; Ruth Ahnert, “Maps Versus Networks,” in *News Networks in Early Modern Europe*, ed. Joad Raymond and Noah Moxham (Brill Open E-Book Collection, 2016), 130–57; Roberto Franzosi, Gianluca De Fazio, and Stefania Vicari, “Ways of Measuring Agency: An Application of Quantitative Narrative Analysis to Lynchings in Georgia (1875–1930),” *Sociological Methodology* 42 (August 2012): 1–42; Shin-Kap Han, “The Other Ride of Paul Revere: The Brokerage Role in the Making of the American Revolution,” *Mobilization: An International Quarterly* 14, no. 2 (2009): 143–62.

ation and editing of historical articles? As the philosopher Bruno Latour has put it, networks can shed light on how a group of people or community is formulated; they reveal the common things and elements that contribute to the formation of a group.⁴ In this chapter, I argue that the production of historical knowledge on Wikipedia is produced through power and hierarchy. By studying multiple Wikipedia pages as networks, I explore the interests of Wikipedia editors, who engage with articles related to history, and their identities as they are promoted within the Wikipedia community. At the same time, by examining the number of edits Wikipedians make and the centrality of Wikipedia editors within a network, I argue that the most experienced editors and some non-human agents define the production of historical knowledge on Wikipedia to a significant extent.

To answer these questions, I have selected three categories related to modern United States history: “History of organizations based in the United States,” “History of science and technology in the United States,” and “LGBT history in the United States”.⁵ All three categories belong to the WikiProject United States, a collaborative project of Wikipedians who aim to improve coverage of topics related to the United States, such as culture, economy, geography, health, history, etc.⁶ Rather than choosing random Wikipedia pages related to modern US history, I have opted to follow the structure and logic of how Wikipedia itself organizes its topic pages by category. Each category consists of multiple pages, forming an organic collective of loosely affiliated pages. I view each category as an actual network of editors. The category “History of organizations based in the United States” includes twenty-two pages, the category “History of science and technology in the United States” has forty-nine pages, and the category “LGBT history in the United States” has seventy-one pages. All three categories represent completely different research interests and fields of history and, thus, I can better detect similarities and differences in Wikipedia editors’ engagement with history.

The topics of the pages of each category vary significantly. The category “History of organizations based in the United States” includes pages that range from

4 Bruno Latour, *Reassembling the Social. An Introduction to Actor-Network-Theory* (Oxford University Press, 2005).

5 For the three categories, see “Category:History of organizations based in the United States,” *Wikipedia*, accessed September 10, 2021, https://en.wikipedia.org/wiki/Category:History_of_organizations_based_in_the_United_States; “Category:History of science and technology in the United States,” *Wikipedia*, accessed September 10, 2021, https://en.wikipedia.org/wiki/Category:History_of_science_and_technology_in_the_United_States; “Category:LGBT history in the United States,” *Wikipedia*, accessed September 10, 2021, https://en.wikipedia.org/wiki/Category:LGBT_history_in_the_United_States

6 For the WikiProject United States, see “Wikipedia:WikiProject United States,” *Wikipedia*, accessed September 10, 2021, https://en.wikipedia.org/wiki/Wikipedia:WikiProject_United_States

the “History of the Industrial Workers of the World” and “History of the Democratic Party” to the “History of the San Francisco Police Department” and “History of the Alcoholics Anonymous”.⁷ The category “History of science and technology in the United States” consists of pages such as “Technological and industrial history of the United States,” “The Machine in the Garden,” and “Electro-Dynamic Light Company”.⁸ The last category has pages that range from “Gay pride” and “LGBT culture in Philadelphia” to “Larry Craig scandal” and “International Pronouns Day”.⁹ So, many of these pages are often loosely connected with each other thematically, but this is just how Wikipedia has chosen to categorize its contents related to modern United States history.

In each category, I have extracted the names of the top twenty editors for each page and the number of edits they have made to each page. I extracted only the names of the top twenty editors because, in the “page statistics” portal, Wikipedia displays only the top twenty editors in the foreground of each page – if someone wants to find out about the rest of the editors, they need to click on “others”.¹⁰ The top editors are identified by the number of edits they have made to each page and should not be confused with the top authors, who are measured by character count. The process for gathering this data was done automatically by using a script, which extracted the details of the top twenty editors for every page as well as the total number of edits made by each one.¹¹ Then, I manually assigned attributes to each Wikipedia editor by visiting their profile pages.¹² That process could not take place automatically, as the profile pages required close reading and interpretation. For example, to understand which topics these editors were interested in on Wikipedia, it would often require a close reading of a few paragraphs that the editors had included within their own profile pages. Therefore, according to the available information on these profile pages, I collected data on each editor’s education, interests, and experience. Specifically, I exam-

7 “Category:History of organizations based in the United States”.

8 “Category:History of science and technology in the United States”.

9 “Category:LGBT history in the United States”.

10 For example, see the “page statistics” of any Wikipedia page: “Science and technology in the United States,” *Wikipedia*, accessed September 10, 2021, https://xtools.wmflabs.org/articleinfo/en.wikipedia.org/Science_and_technology_in_the_United_States#top-editors

11 By creating a script written in the Python programming language, I automatically scraped the names of the top twenty editors and the number of their edits from all pages of each Wikipedia category. For example, I applied the script to the category “History of organizations based in the United States” and it imported the names of the relevant top twenty editors and the number of their edits from each Wikipedia page that belongs to this category.

12 For example, see the profile page of the user “Settler,” “User:Settler,” *Wikipedia*, accessed September 10, 2021, <https://en.wikipedia.org/wiki/User:Settler>

ined if they had any educational background, such as an undergraduate or graduate degree in history, if they were interested in history as a subject of knowledge, if they were interested in the general topic they had chosen to edit, such as science, technology, LGBTQ, or the specific topic of the article itself, and if they were experienced users. As Wikipedia does not include any information on what distinguishes an experienced user from an inexperienced one, I decided to set my own criteria for what constitutes an experienced user: this was either the editing of at least ten different articles, or having been the recipient of any barnstars or other editing awards by the Wikipedia community. In addition, I extracted information about editors' experiences and interests from the "view history" portal of each profile page, which details the past activities of editors on Wikipedia.¹³ For example, if an editor was making constant contributions to articles related to history, they were marked as experienced and interested in history editors for the purposes of the network analysis.

On their profile pages, some editors mentioned the reasons why they got involved on Wikipedia, their studies, interests, their awards for their contributions, how long they had been participating on Wikipedia, and even some more personal information. However, many editors did not include any information about their lives, activities, or interests, even if they had contributed to multiple pages. Some of them were also banned from Wikipedia, so there was no available information about their role or activities within the community. Wikipedia defines its bans as "formal prohibition from editing some or all Wikipedia pages, or a formal prohibition from making certain types of edits on Wikipedia pages".¹⁴ Bans can have a specific or unspecified duration.¹⁵ On the one hand, this lack of information prevents us from having a complete picture of all editors' characteristics. On the other hand, the available data reveals the frequency with which Wikipedia editors appear within a certain category, their position within this category, the number of edits they have made, and their relation to history and the broader thematic topic they have chosen to edit. Once I had collected all the relevant data and assigned attributes to the various editors (interest in history, interest in the specific topic, education in history, experience, etc.), I imported this data to Gephi, a visualization and exploration software for graphs and networks.¹⁶

¹³ For example, see "User:Settler: Revision history," *Wikipedia*, accessed September 10, 2021, <https://en.wikipedia.org/w/index.php?title=User:Settler&action=history>

¹⁴ "Wikipedia:Banning policy," *Wikipedia*, accessed September 10, 2021, https://en.wikipedia.org/wiki/Wikipedia:Banning_policy

¹⁵ *Ibid.*

¹⁶ On Gephi and its usage for network analysis, see Ahnert, et al., *The Network Turn*, 64.

In the Wikipedia community, there are editors and pages, while in networks there are nodes and edges. A network is a “set of relationships between objects or entities”.¹⁷ Nodes are groups of objects or entities, and their relationships are edges.¹⁸ In this study, Wikipedia editors will be the nodes and the connections between the editors, who contribute to the creation and development of a Wikipedia page, are the edges. In other words, the edges represent a connection of one editor with another editor, and this connection is established by the fact they both edited the same page. Measuring the connections between Wikipedia editors helps us understand the number of edits these editors have made to each page, the importance of specific individuals, and lets us identify who are the most active and well-connected editors within a given network. As Ruth Ahnert and Sebastien Ahnert explain in their study on networks: “A network is a collection of links, which can be combined into a myriad of possible paths. The measurement of these paths is a crucial way of establishing the ranked importance of the people in that network”.¹⁹ Therefore, network analysis is also political, it reveals the dynamics of power.²⁰ In the case study of Wikipedia, some experienced editors are responsible for making the most contributions to Wikipedia articles, and some non-human agents have the highest numbers of connections within a Wikipedia network, and, therefore, determine the production of historical knowledge.

Exploring Wikipedia editors’ interests and identities

Each Wikipedia page constitutes a network with nodes and edges and each page is part of a broader network of a Wikipedia category. In this study, I have named each category as a category-network and the involved Wikipedia pages as pages-networks. The aim of this section is to explore the different factors and various interests that motivate Wikipedia editors to engage with the production of historical knowledge on Wikipedia. By examining three different historical categories on Wikipedia and multiple Wikipedia pages, I was able to identify the types of interests held by Wikipedians, who write historical articles, and which encourage

17 Ahnert, “Maps Versus Networks,” 131.

18 Ibid. On networks and their characteristics, see Franco Moretti, *Distant Reading* (London: Verso, 2013), 213; Scott B. Weingart, “Demystifying Networks,” *The Scottbot Irregular*, accessed December 1, 2021, <http://www.scottbot.net/HIAL/index.html@p=6279.html>; Elijah Meeks, “More Networks in the Humanities or Did books have DNA?” *Digital Humanities Specialist*, accessed December 1, 2021, <https://dhs.stanford.edu/visualization/more-networks/>

19 Ahnert and Ahnert, “Protestant Letter Networks,” 12.

20 Ahnert, et al., *The Network Turn*, 40.

them to produce historical knowledge. In turn, these interests offer valuable insight into the respective identities of the Wikipedia editors in question, and how they choose to promote themselves within the Wikipedia community.²¹ The main argument of this section is that these Wikipedia users, who edit articles related to history, identify themselves more with the specific topic they edit rather than with history as a general field or category of interest. At the same time, any formal education in history, which these Wikipedians may or may not have received, does not play any role in their engagement with articles related to history.

Figure 27 (see below) is a visualized network of the Wikipedia category “History of organizations based in the United States”. This category includes twenty-two pages-networks. The visualization of networks is always the result of specific choices and assumptions. The network in Figure 27 emphasizes the nodes-editors with the highest number of edits within the category. The bigger the node, the higher the number of edits the node has made. Same for the color. The lighter the blue of the node, the higher the number of edits the node has made, and vice versa. As can be seen in the network, not all editors make an equal number of contributions within the category. There are significant differences in the number of edits Wikipedians make. Even if we zoom in on the network and scrutinize a particular page-network more closely (Figures 28 and 29), we notice that some of the involved users have made a higher proportion of contributions than other editors.²² In this section, I identify the editors with the highest numbers of edits, their interests and identities that encourage them to produce historical knowledge, and finally their centrality within each category-network.

By looking again at the broader category-network (Figure 27), three editors are the top editors of the category. The users “Rjensen,” “Settler,” and “Richard Myers” appear to be the editors with the most contributions, with more than two hundred edits each. Then there are users “AHC300,” “Hourick,” “Chris Light,” and “207.232.97.13” with a lower number of edits, more than one hundred each. Why do these users make the most edits? Firstly, it is to do with the topics they edit. “Rjensen,” “Settler,” and “AHC300” are involved with the creation and editing of the pages “History of the Democratic Party” and “History of the Republican Party”. “Richard Myers” contributes to the “History of the Industrial Workers of the World,” “Hourick” to the “History of the Houston Police Department,” “Chris Light” to the “History of the National Park Service,” and “207.232.97.13” to the “History of Alcoholics Anonymous”. All those pages represent very popular topics,

²¹ On how identities work in online communities, see Bruckman, *Should You Believe Wikipedia?*, 118–59.

²² The pages-networks are distinguishable because they have a high number of exclusive editors despite the smaller number of editors shared with other clusters.

which are central to modern United States history, especially the “History of the Democratic Party” and the “History of the Republican Party”. Furthermore, those topics are broader than, for example, the “History of the National Register of Historic Places” or the “History of the Texas Ranger Division”. The more popular and broader the topics, the more edits they attract. This also relates to the process of consensus-making. The most popular and central topics within a category tend to attract more



Figure 27: History of organizations based in the United States; in this and the following figures: color – lighter color means more edits and darker color means fewer edits; node size – number of edits.

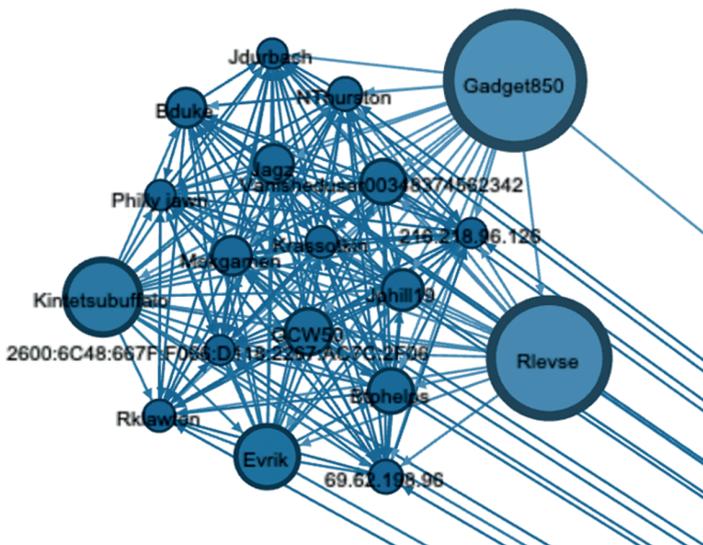


Figure 28: History of the Boy Scouts of America/History of organizations based in the United States.

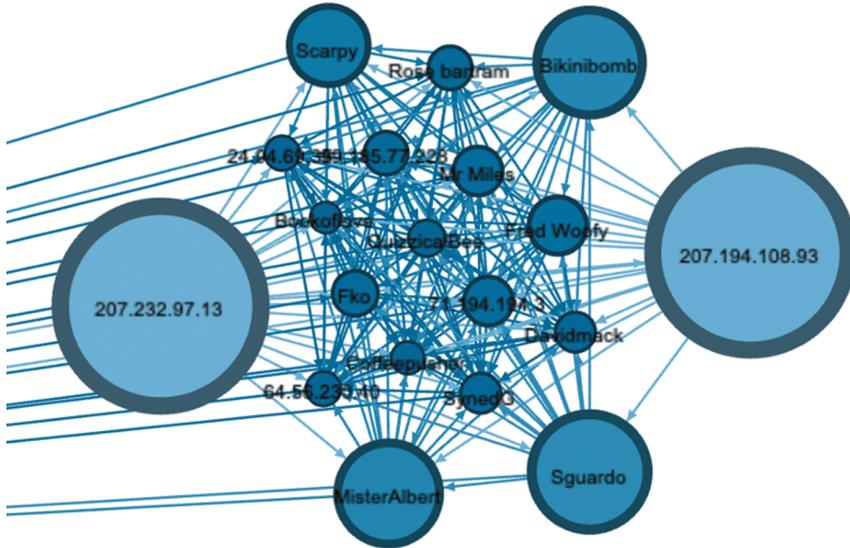


Figure 29: History of Alcoholics Anonymous/History of organizations based in the United States.

editors, who also present different perspectives and points of view. Therefore, other editors need to constantly intervene in order to ensure appropriate edits are made on a given page. However, this cannot be detected and visualized in a network.

Another key consideration has to do with the individual characteristics of Wikipedia editors. In Figure 30, we see how many editors are interested in the topic they edit. Those displayed in a light blue color are interested in the topic that they edit, those in orange do not include any information about their interests, and those displayed in pink are not interested in the topic. Almost sixty-five percent of editors do not include any information about their interests; thirty-three percent are interested, and two percent are not. Even though most users do not clarify their interests within the Wikipedia community, it is evident that those editors responsible for high numbers of edits demonstrate a clear interest in the topic to which they have chosen to contribute. As shown in Figure 30, “Rjensen,” “Settler,” “Hourick,” “Richard Myers,” “Chris Light,” and “207.232.97.13” are all interested in the relevant topic. Only “AHC300” is displayed in orange, as this user does not disclose any information about their interests.

This pattern that the editors with the highest number of edits are also interested in the topic they choose to edit, however, does not only characterize the editors with the most contributions. If we zoom in on a smaller page-network of the category, we can detect the same pattern. For example, in Figure 31, we see the page “History of the New York City Bar Association”. The editors with the most



Figure 30: History of organizations based in the United States.

edits, “Pattonnh” and “66.104.196.194,” have an interest in this topic. The other editors in the network do not include information in their profiles, and one user is not interested at all. The same happens in the page “History of the San Francisco Police Department,” in Figure 32. “Hank Chapot,” the top editor of the page, is interested in the topic, the rest of the editors are either interested in the topic or they do not disclose any information about their interests. The same pattern characterizes even smaller networks, in which the difference between the top editor and the rest of the editors is not that high, as in Figures 33 and 34. Therefore, the interest in the topic determines the number of edits that editors make. Of course, the more popular the theme of a page, the higher the number of edits the page attracts. But even on pages that relate to more specialized thematic areas, the editors with the most edits appear to be interested in that specific topic. The top editors are not only interested in what they edit but they also keep an active profile page, in which they refer to their research interests within the Wikipedia community.

Even if we look at the category of “History of Science and Technology in the United States,” which has a completely different thematic area to the previous examples, the same results appear. Five editors are the top editors of the category, and four out of the five disclose their interest in the topic (Figures 35 and 36). When we zoom in on smaller pages-networks (Figures 37 and 38), we can again see the same pattern. On most pages, the editor with the highest number of contributions is interested in the topic they edit. The same results appear when we look at the category “LGBT history in the United States” (Figure 39). In this network, eight editors have more than one hundred edits, as the category of LGBT history has more pages than the other two categories. Also, this category includes topics that can be regarded as more controversial than the articles in the

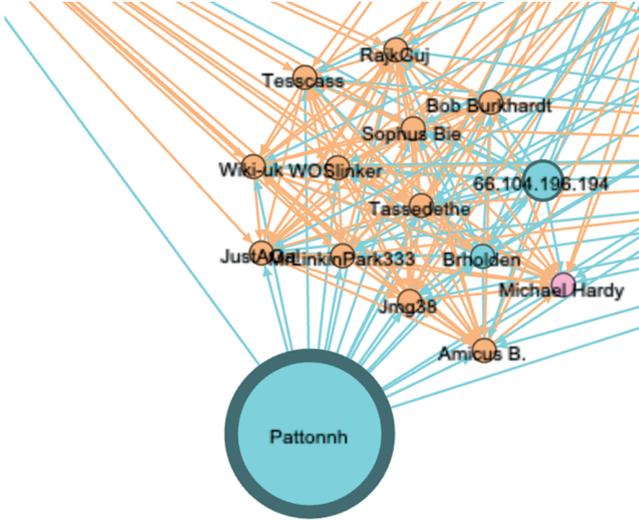


Figure 31: History of the New York City Bar Association/History of organizations based in the United States; in this and the following figures: color – light blue means interested in the topic they edit, orange color means no available information, and pink means not interested in the topic; node size – number of edits.

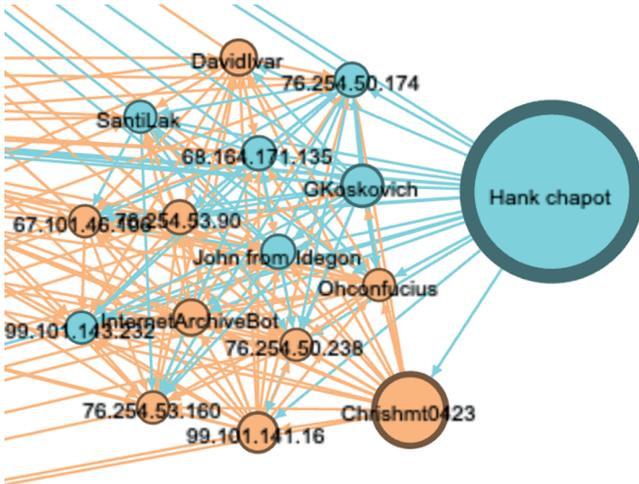


Figure 32: History of the San Francisco Police Department/History of organizations based in the United States.

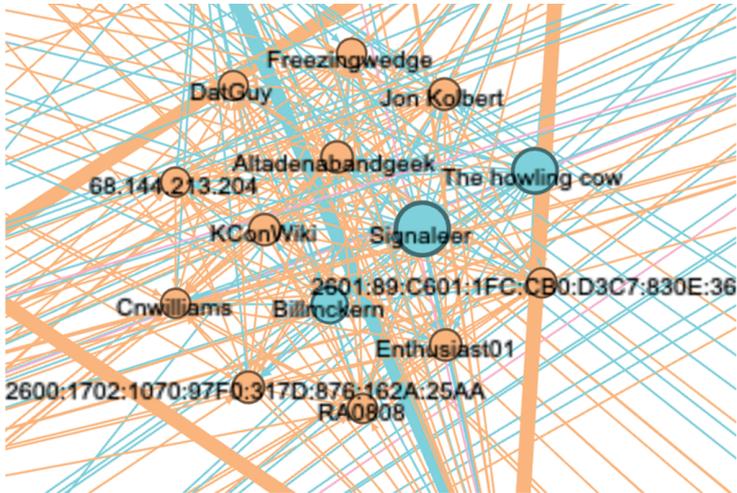


Figure 33: History of the United States Army National Guard/History of organizations based in the United States.

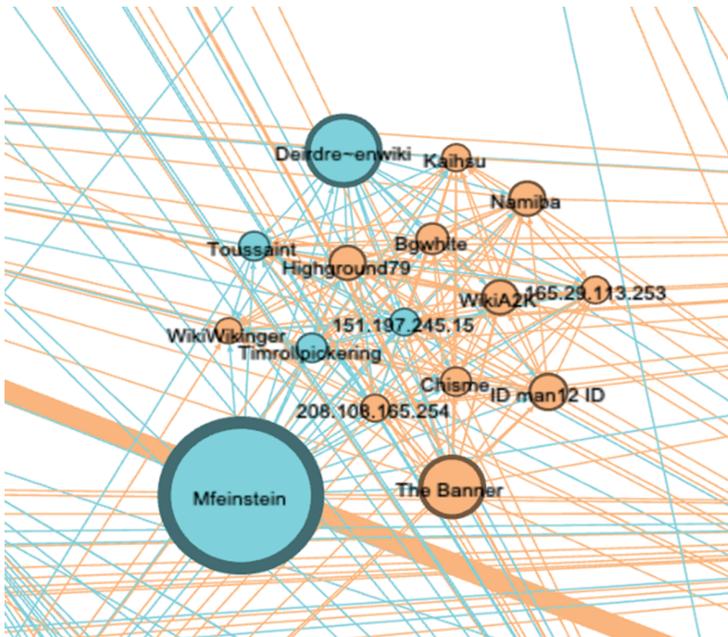


Figure 34: History of the Green Party of the United States/History of organizations based in the United States.

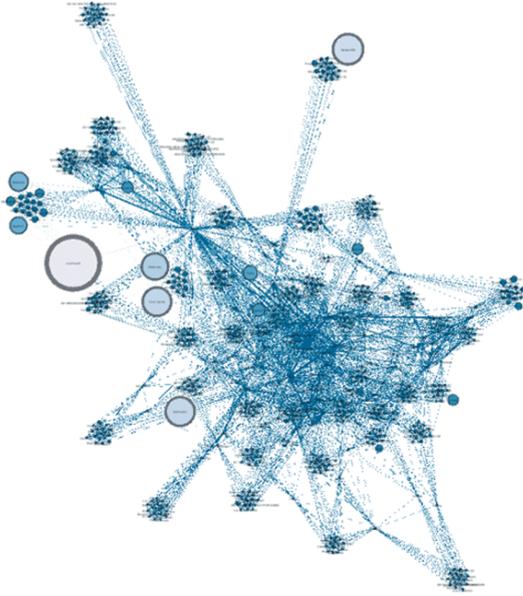


Figure 35: History of Science and Technology in the United States; color – lighter color means more edits and darker color means fewer edits; node size – number of edits.

other two categories. The more controversial the pages, the higher the editors' level of engagement, as they try to reach a consensus. In Figure 40, we detect the same pattern regarding the interest of editors in the topic they edit. All the top editors of the category-network are interested in the relevant topic. If we zoom in and examine the pages-networks more closely, we can see the same results in most cases (Figures 41–43).

It is evident that most top editors in each category are highly interested in the topic to which they contribute. Overall, in all three examined categories-networks, thirty-five percent of editors are interested in the topic they edit, and almost all the editors with the highest number of contributions are interested in the topic. It is the topic itself which drives and mobilizes Wikipedia editors to engage with the creation and editing of Wikipedia pages related to history. What about other factors, though, such as having an interest in history or an educational background related to history? Do those factors encourage Wikipedia users to contribute to Wikipedia?

Figure 44 displays users who are interested in history. The editors who are interested in history are displayed in light blue, those who are not interested are displayed in pink, and those who do not include any information about their interest in history are displayed in orange.

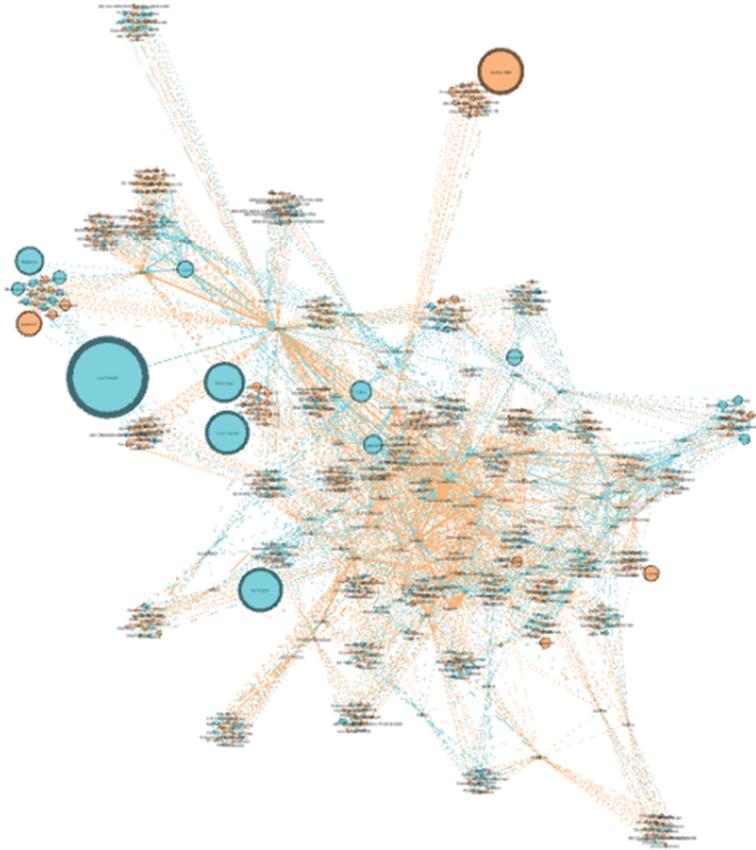


Figure 36: History of Science and Technology in the United States; in this and the following figures: color – light blue means interested in the topic they edit, orange color means no available information, and pink means not interested in the topic; node size – number of edits.

Less than half of the editors, who are interested in this topic, are interested in history. Specifically, in the category “History of organizations based in the United States,” thirteen percent of the editors are interested in history, while thirty-three percent are interested in the topic (Figure 44). Of the seven top editors in this category, three are interested in history. The rest just mention that they are interested in the topic itself. Similar results appear in the other two categories. In the category “History of science and technology,” eighteen percent of editors are interested in history, while thirty-five percent are interested in the topic (Figure 45). Of the five top editors, three are interested in history. In the category “LGBT history in the United States,” eighteen percent of editors are interested in history and thirty-eight

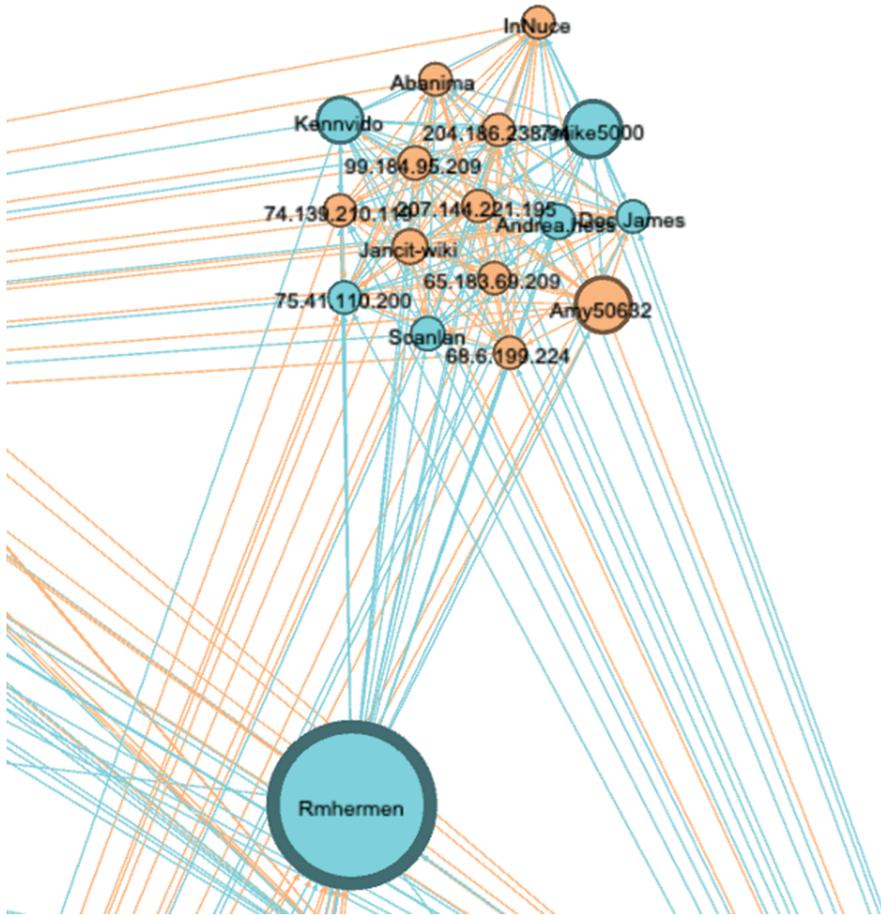


Figure 37: West Nile virus in the United States/History of Science and Technology in the United States.

percent are interested in the topic (Figure 46). Of the eight top editors in this category, two are interested in history. The users identify themselves more with the topic itself rather than with history as a field or category of interest.

This point is interesting as it reveals how the concept of “history” is perceived within the Wikipedia community. The fact that many Wikipedia users create and edit historical pages, while they are interested only in the topic itself rather than history more generally, emphasizes the need for a retheorization of “history” within the context of Wikipedia as a digital and public space. Wikipedians’ identification with individual topics and history more broadly is reminiscent of Rose-

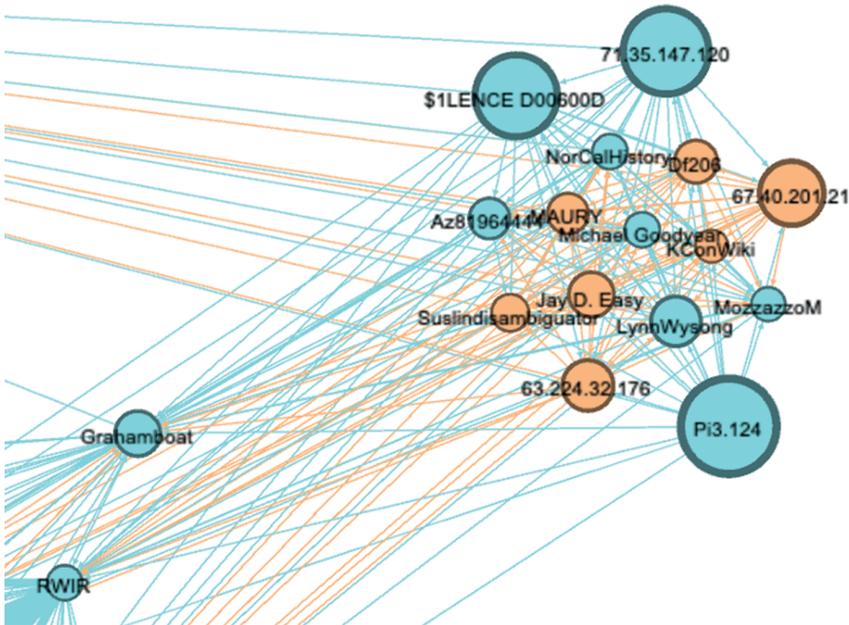


Figure 38: United States Exploring Expedition/History of Science and Technology in the United States.

nzweig and Thelen’s work, *The Presence of the Past*, in which “the past” was present in the interviews conducted, but “history” was not – or, at least, not how it has been defined in the academic textbooks.²³ Interviewees tended to associate “history” with a boring school class, whereas “the past” was seen to contain a volume of information that could help them with their lives. Wikipedia’s distinction between topic and history is similar to Hayden White’s distinction between the “historical past” and the “practical past”.²⁴ On Wikipedia, the topic serves as the “practical past,” which has a political and practical use in the present and encourages most Wikipedia editors to create and edit historical articles. History becomes the “historical past,” a more scientific and distant past, which appears to motivate fewer editors to engage with Wikipedia’s historical contents.

However, as we have seen in the previous chapters both of these pasts coexist within the Wikipedia community and define users’ engagement and the production of historical knowledge. Even if history is not the most significant factor that

²³ Rosenzweig and Thelen, *The Presence of the Past*, 9.

²⁴ White, “The Practical Past,” 10–19; White, “Politics, History and the Practical Past,” 127–34.

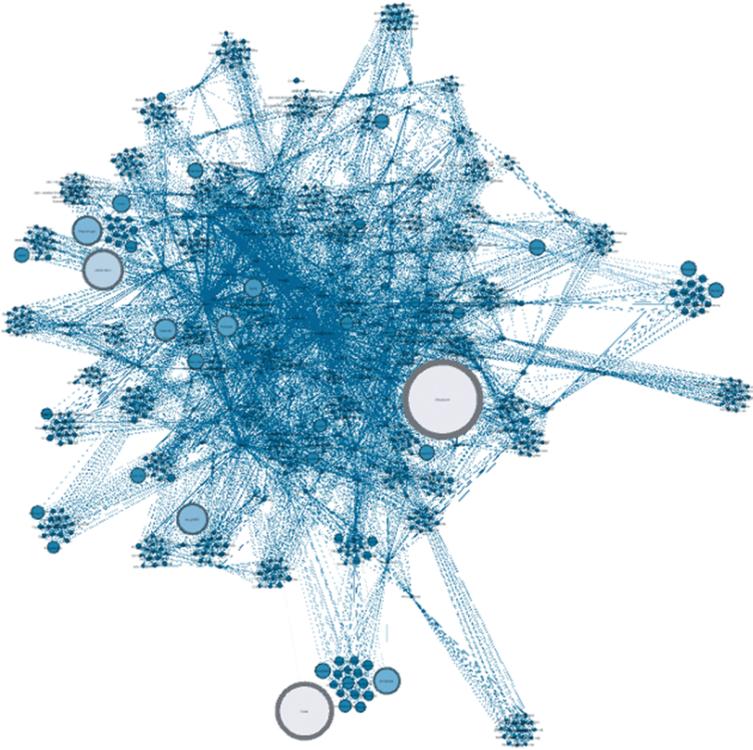


Figure 39: LGBT history in the United States; color – lighter color means more edits and darker color means fewer edits; node size – number of edits.

encourages Wikipedians to create and edit articles, it mobilizes a great number of editors to an important extent. By looking closer at the pages-networks, we notice that in most cases there are few editors who are interested in history, but most of them have a significant number of contributions. Of course, there are pages-networks in which no one is interested in history; however, in most cases, there will be at least one editor interested in history. These patterns appear in all three categories (Figures 47–56).

On their profile pages, some editors include information about their educational background. In each category studied here, only two percent of Wikipedia editors have references to their educational backgrounds. Specifically, in the category “History of organizations based in the United States,” two percent of editors claim that they have studied history either at undergraduate or graduate level (displayed in pink), seven percent have studied a subject or field other than his-

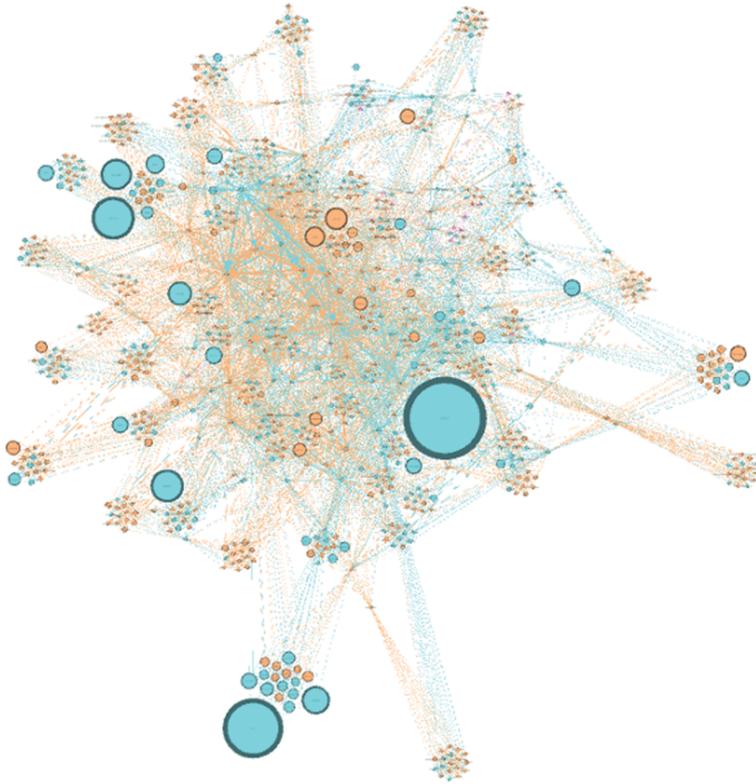


Figure 40: LGBT history in the United States; in this and the following figures: color – light blue means interested in the topic they edit, orange color means no available information, and pink means not interested in the topic; node size – number of edits.

tory (light blue), and ninety percent do not include any information about their education (orange) (Figure 57). In the category “History of Science and Technology in the United States,” less than one percent of editors have any education in history, six percent have an education in a different field, and ninety-three percent do not include information about their educational background (Figure 58). In “LGBT history in the United States,” one percent of users have studied history, three percent has studied any other field, and ninety-six percent have not disclosed any information about their studies (Figure 59).

Most Wikipedians do not include any information about their degrees or the studies they may have completed. Education does not play any role in users’ engagement with history, and even more significantly, education does not define the identity of most Wikipedians. Even the editors who have studied in a different

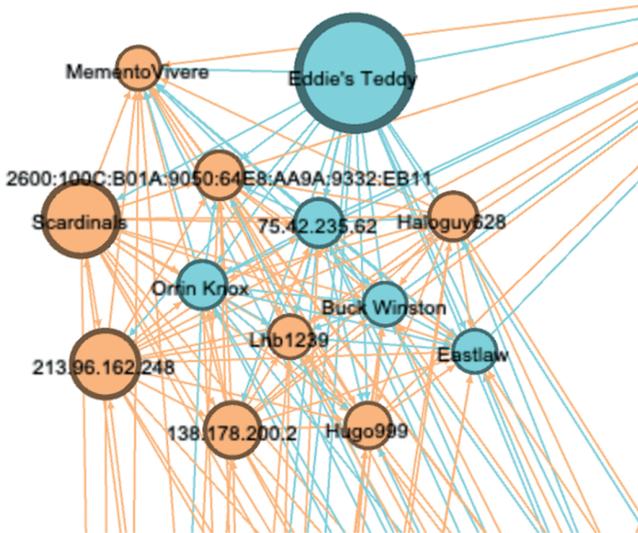


Figure 41: United States Navy dog handler hazing scandal/LGBT history in the United States.

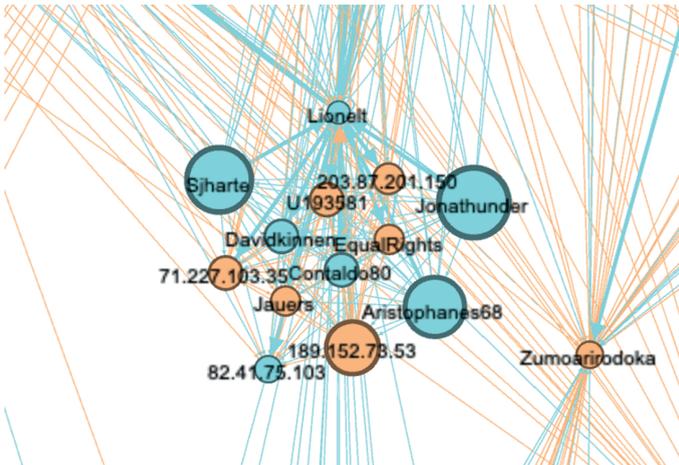


Figure 42: Metropolitan Community Church/LGBT history in the United States.

field number more than those who have studied history. This is the case in all three categories. These fields might be close to history, such as literature, political science, sociology, anthropology, but still, even in those cases, studies in history do not appear to shape users' participation on Wikipedia. It is the interest Wikiped-

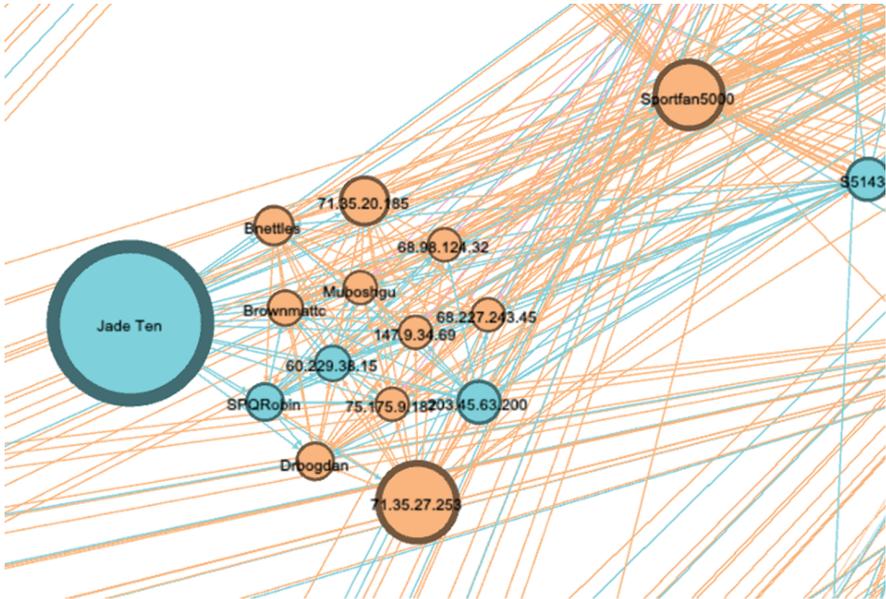


Figure 43: Arizona SB 1062/LGBT history in the United States.

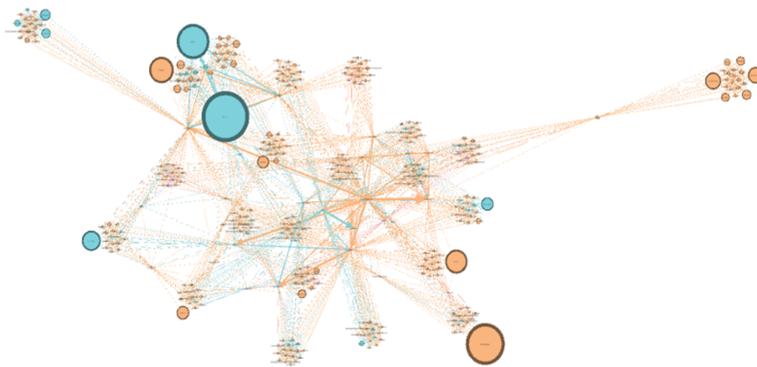


Figure 44: History of organizations based in the United States.

dians have in the topic they edit that defines users' engagement with history and shapes their virtual identities on Wikipedia. The interest in history is also significant, as it determines users' participation within the Wikipedia community to a large extent. Education appears to be the least relevant factor when it comes to encouraging Wikipedians to write about history. While their educational background may be fixed, the identities of Wikipedia users are fluid; they are deter-

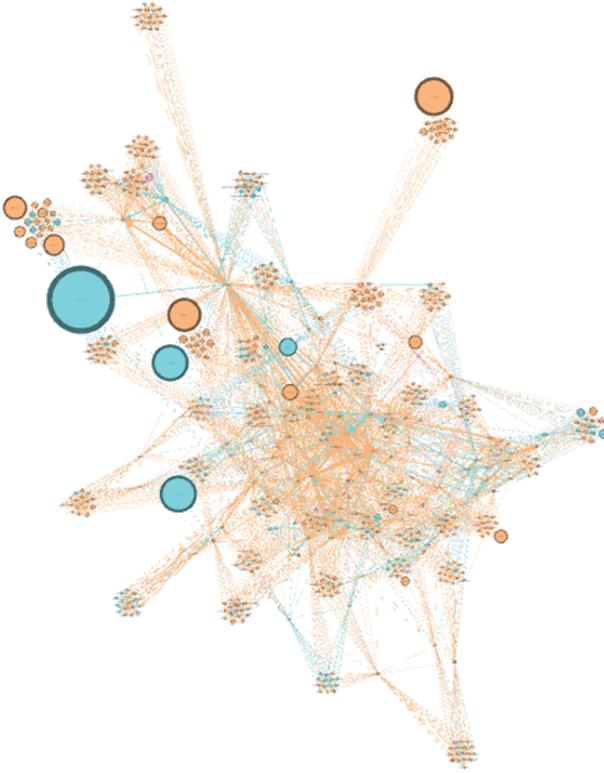


Figure 45: History of science and technology in the United States.

mined by the users themselves and not by the studies they have previously completed or the degrees they have earned. Wikipedia constitutes a digital and public space in which editors can display and demonstrate their interests and not just the skills or knowledge they have obtained from their educational training. The interests of Wikipedians can constantly change and, therefore, this allows users to present themselves as they wish within the Wikipedia community, and, even more importantly, to become experts in the areas they are passionate about.

Bots, veterans, and newbies

Interest in a given topic, and to a lesser extent in history more generally, encourages most Wikipedia users to create and edit Wikipedia articles. However, the actions of some agents within a category-network remain unknown. What happens to the users who are neither interested in the topic they edit nor in history? Why

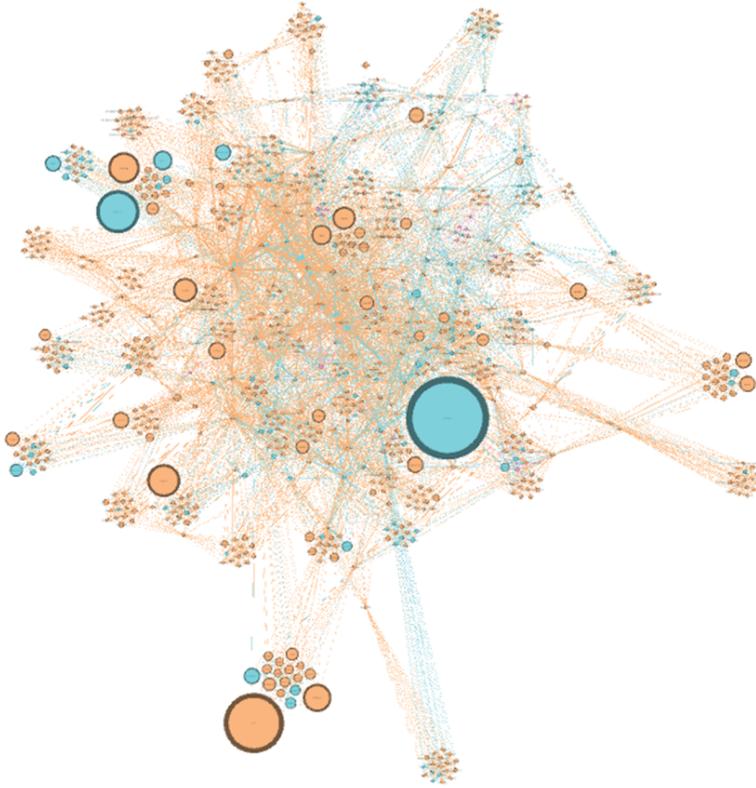


Figure 46: LGBT history in the United States.

do they edit Wikipedia articles about history? How do they get engaged with the production of historical knowledge on Wikipedia? To answer these questions, we need to think about the digital setup of Wikipedia and the existence of non-human agents within the wider networks of engagement and interactions. In this section, by looking at the centrality of Wikipedia editors within a category-network, I reveal the structures of power that determine the different roles of Wikipedia editors and define the production of historical knowledge on the site.²⁵

In the previous section, the size of each node was based on the number of their contributions in the three categories-networks. Most of these nodes (the relevant editors) are interested either in the topic itself or in history more generally. But what about the rest? Let us look at the networks again but this time visualize

²⁵ For a broader analysis of the social roles of people in online communities, see Bruckman, *Should you Believe Wikipedia?*, 24–7.

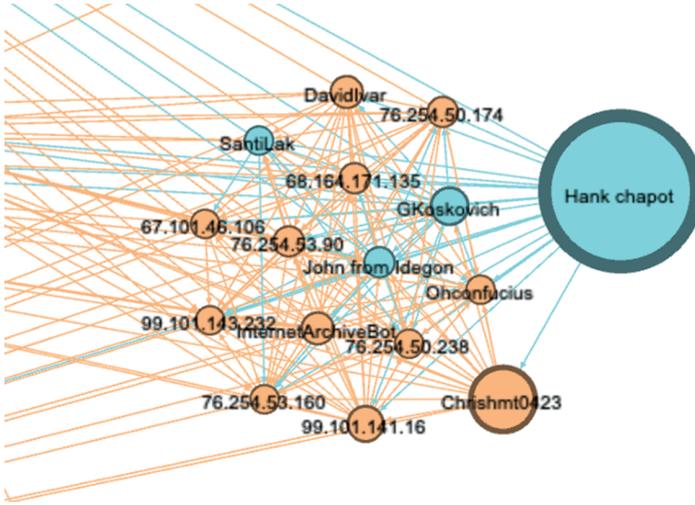


Figure 47: History of the San Francisco Police Department/History of organizations based in the United States.

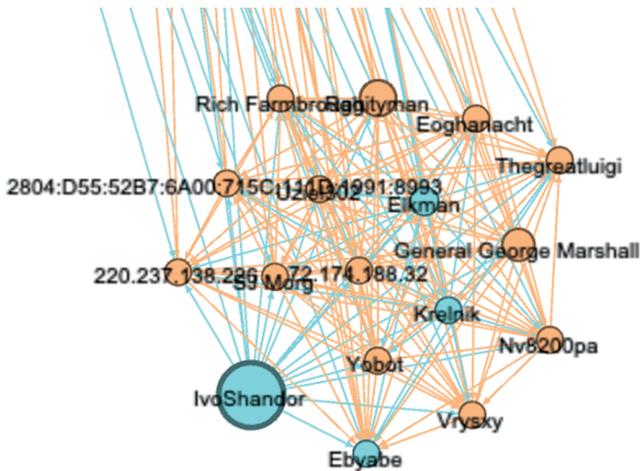


Figure 48: History of the National Register of Historic Places/History of organizations based in the United States.

the nodes according to their centrality within each category-network, namely, according to the number of editors' connections (edges). Different results appear. In the category-network "History of organizations based in the United States," the most central editors are different from those with a high number of contribu-

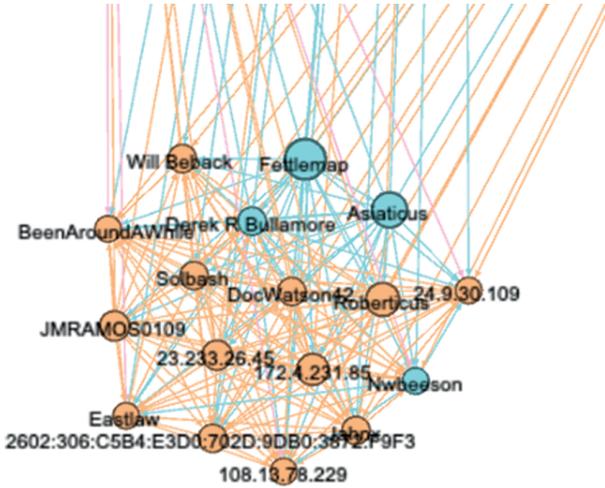


Figure 49: History of the Los Angeles Police Department/History of organizations based in the United States.

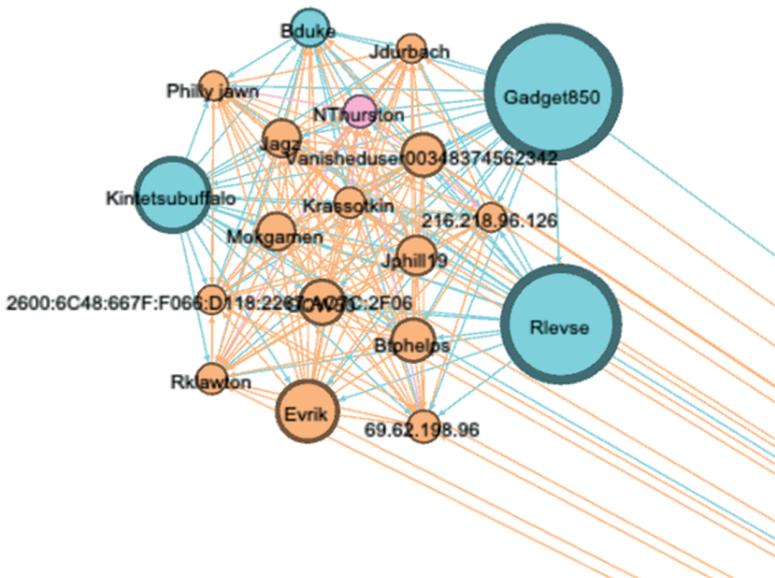


Figure 50: History of the Boy Scouts of America/History of organizations based in the United States.

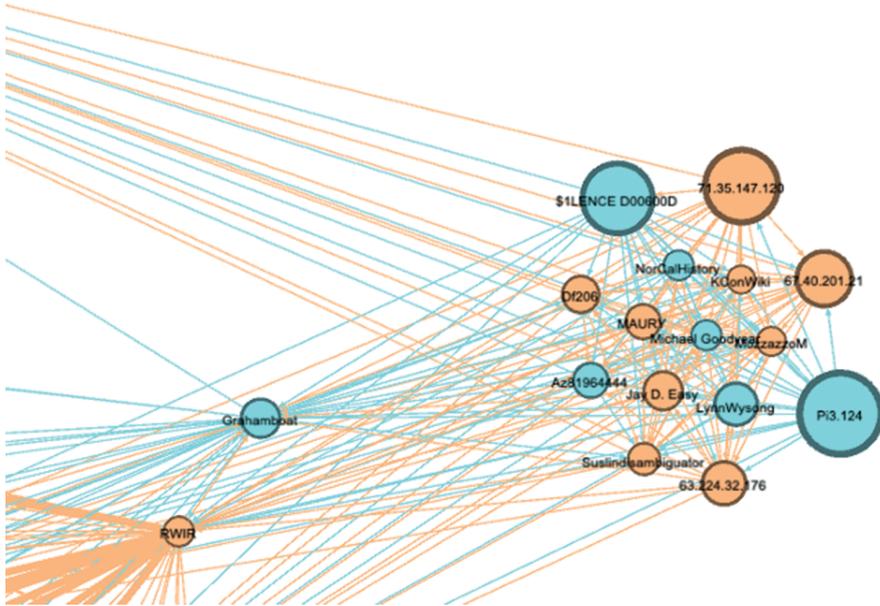


Figure 51: United States Exploring Expedition/History of Science and Technology in the United States.

tions. In Figure 60, the editors with the highest number of connections (edges) are displayed with a larger node and in a light pink color. The editors with more connections are the editors who appear on more pages. In other words, the editors with more connections are those who have edited more pages. Those editors should not be confused with the editors with higher numbers of edits.

By looking at Figure 59 more closely, we notice that the editors with the most connections are: the “InternetArchiveBot,” the “ClueBot NG,” the “Cydebot,” “Tim!,” “Monkbot,” and “Srich32977”. Then we have “Illegitimate Barrister,” “AnomieBOT,” “KolbertBot,” “GreenC bot,” and “Hmains”. It is more than obvious that most of those editors have a common characteristic. They are bots. According to Wikipedia, a bot “is an automated tool that carries out repetitive and mundane tasks to maintain the 54,264,985 pages of the English Wikipedia”.²⁶ The communication scholar Randall Livingstone has defined Wikipedia bots as “a program or script that carries

²⁶ “Wikipedia:Bots,” *Wikipedia*, accessed September 10, 2021, <https://en.wikipedia.org/wiki/Wikipedia:Bots> On the history of Wikipedia bots, see “Wikipedia:History of Wikipedia bots,” *Wikipedia*, accessed September 10, 2021, https://en.wikipedia.org/wiki/Wikipedia:History_of_Wikipedia_bots

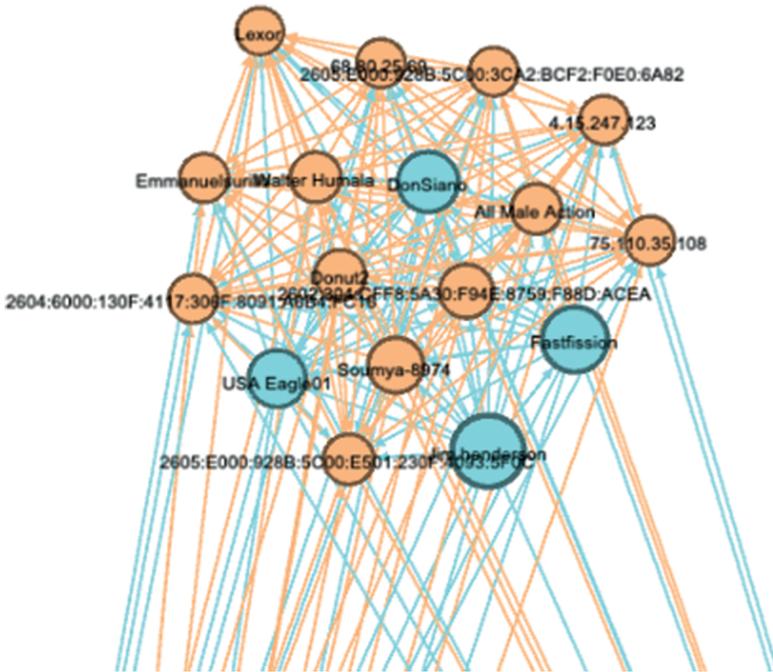


Figure 52: Science and technology in the United States/History of Science and Technology in the United States.

out an often tedious or repetitive tasks for its creator”.²⁷ They are created in the python programming language by Wikipedia users and some of the bots have the same user namespace and talk pages on Wikipedia.²⁸ Some bot operators use photographs and descriptions to anthropomorphize their bots.²⁹ The first bots appeared on Wikipedia in late 2001, but in 2002 bots started to have an active role on Wikipedia.³⁰ Wikipedia has established bot policies to control its bots.³¹ As José van Dijck has explained, there are two types of bots, the editing or coauthoring bots and non-

²⁷ Randall Livingstone, “Immaterial Editors: Bots and Bot Policies Across Global Wikipedia,” in *Global Wikipedia: International and Cross-Cultural Issues in Online Collaboration*, 7.

²⁸ *Ibid.*, 10.

²⁹ *Ibid.*

³⁰ *Ibid.*, 12. van Dijck, *The Culture of Connectivity*, 137.

³¹ Livingstone, “Immaterial Editors,” 12 and 17. On the bot policy of the English Wikipedia, see “Wikipedia:Bot policy,” *Wikipedia*, accessed September 10, 2021, https://en.wikipedia.org/wiki/Wikipedia:Bot_policy



Figure 53: The Machine in the Garden/History of Science and Technology in the United States.

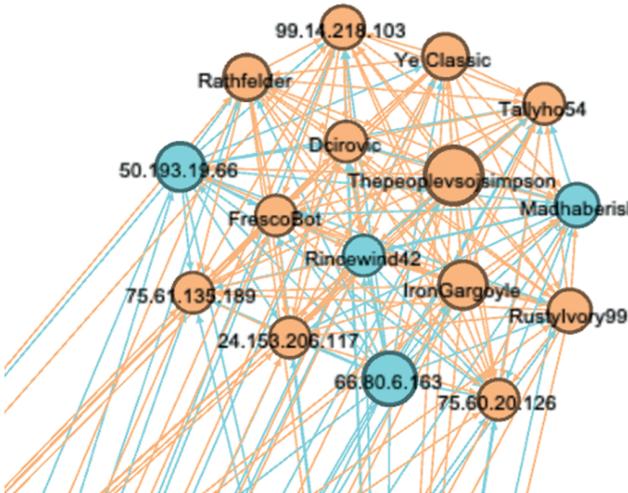


Figure 54: Horizon Services/LGBT history in the United States.

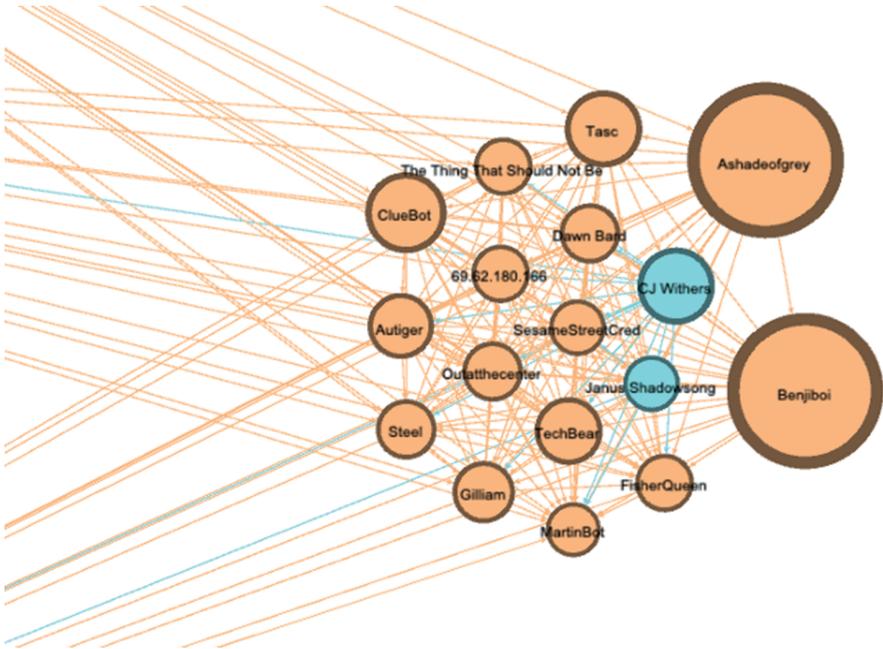


Figure 55: Gay pride/LGBT history in the United State.

editing or administrative bots.³² The administrative bots are very common on Wikipedia. They are responsible for performing policing tasks, such as blocking spam and detecting vandalism, etc. The coauthoring bots are tasked with creating and writing articles about specific topics, but they are not very common on Wikipedia.³³

Even though bots have their own user pages, they do not include information about their interests in particular topics, or in history, or their education.³⁴ Thus, they have been depicted as unknown actors in the networks above. They do not make many edits, only a few to each page, but they are the most central actors within a category-network. They appear on most pages. In Figures 61 and 62, we can see that the more central editors are bots and, therefore, they are displayed as unknown actors (orange) and not as interested in the topic or in history (light blue). Only “Tim!,” “Illegitimate Barrister,” and “Hmains,” who are human actors, are displayed as interested in history and/or the topic.

³² van Dijck, *The Culture of Connectivity*, 138.

³³ *Ibid.*, 138–9.

³⁴ *Ibid.*, 137.



Figure 56: Mariel boatlift/LGBT history in the United States.

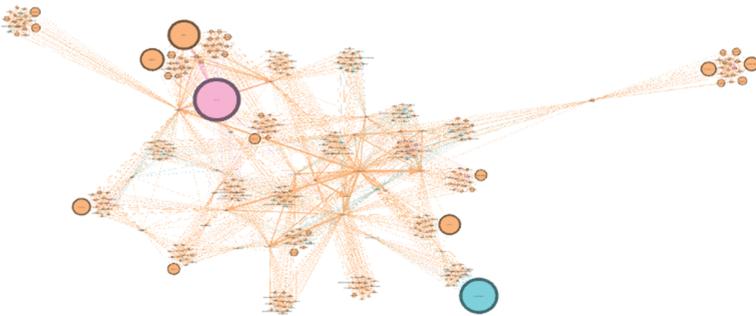


Figure 57: History of organizations based in the United States.

In the category “History of Science and Technology in the United States,” the most central editors are “Monkbot,” “Citation bot,” “ClueBot NG,” “InternetArchiveBot,” “Rjwilmsi,” and “RjwilmsiBot”. Then there is “SmackBot,” “Cydebot,” “Bender the Bot,” “Yobot,” and “Hmains” (Figure 63). Again, the most central editors are bots, and these are effectively the same bots that appear in the previous category, “History of organizations based in the United States”. Only two actors are human, “Rjwilmsi”



Figure 58: History of Science and Technology in the United States.

and “Hmains”. The latter also appear in the previous category. In Figures 64 and 65, we can see that the bots are marked as unknowns in terms of their interest in either the topic or in history, and that “Rjwilmsi” and “Hmains” are interested in both history and technology.

The same results appear in the category “LGBT history in the United States” (Figure 66). The more central actors of the network are “InternetArchiveBot,” “Monkbot,” “Bender the bot,” “AnomieBOT,” “SmackBot,” and “Citation bot.” “Treker,” “GreenC bot,” “Varnent,” “Wikignome0529,” “Bmclaughlin9,” and “Bearcat” have slightly fewer connections. Again, most of these are bots. In Figures 67 and 68, these bots are displayed as unknowns and colored orange. The central human-actors of the network are interested either in the topic or in history. Specifically, “Treker” and “Varnent” are interested both in the topic and history, “Bearcat” and “Wikignome0529” are interested only in the topic, while “Bmclaughlin9” does not include adequate information on their profile page. It is interesting that the user “Hmains,” who appears in the other two categories-networks, is also a central editor in this category-network.

By looking at the centrality of the editors within a category, we can detect two main points. First, it is impossible to truly understand how Wikipedians pro-

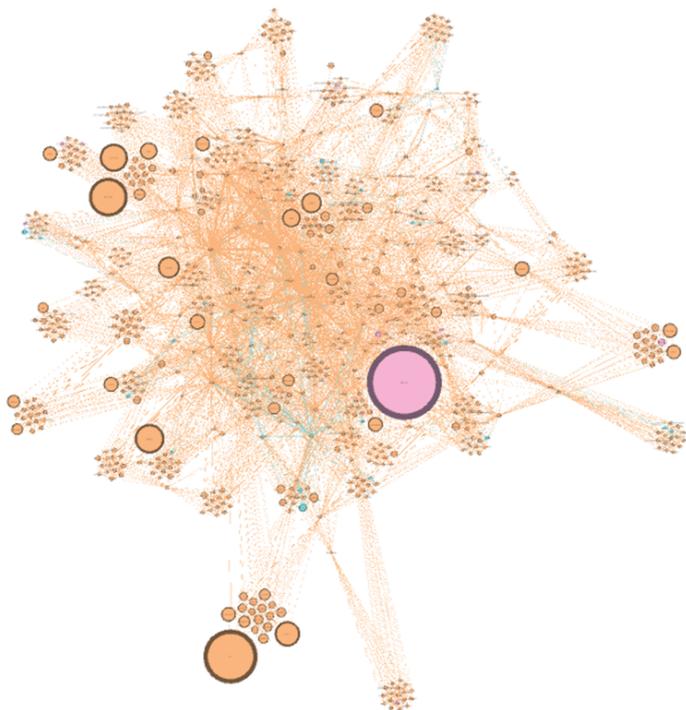


Figure 59: LGBT history in the United States.

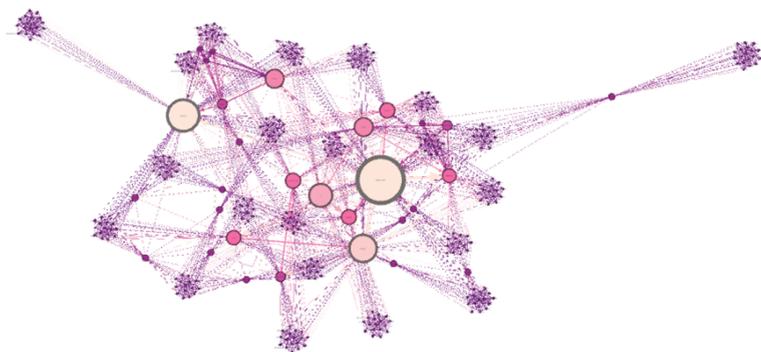


Figure 60: History of organizations based in the United States; color – lighter color means higher number of connections (edges) and darker color means less connections; node size – network centrality.

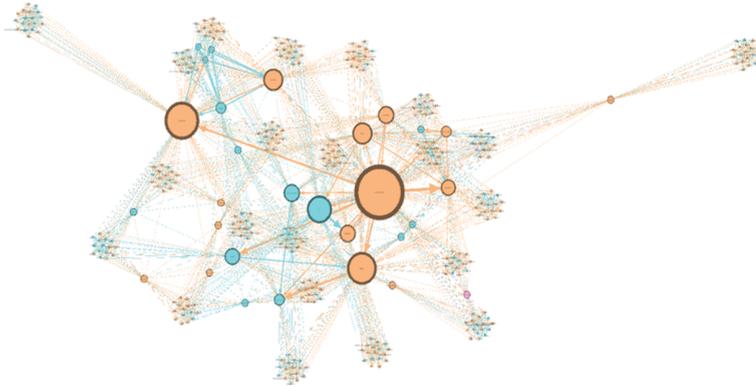


Figure 61: History of organizations based in the United States; color – light blue means interested in the topic they edit, orange color means no available information, and pink means not interested in the topic; node size – network centrality.

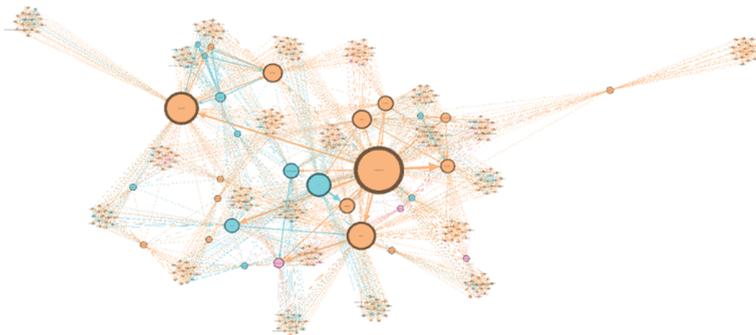


Figure 62: History of organizations based in the United States; color – light blue means interested in history, orange color means no available information, and pink means not interested in history; node size – network centrality.

duce historical knowledge without a consideration of the role played by bots.³⁵ The networks shown above make it clear that bots are the most central actors in a network, as they have the highest number of connections. As Nathaniel Tkacz has pointed out, Wikipedia cannot be understood without its materiality, which is important in terms of any organizational apparatus and includes its servers, software and code, operating systems, web browsers, computers, devices, screens,

³⁵ On the role of bots in the social media universe, see Tony Veale and Mike Cook, *Twitterbots. Making Machines that Make Meaning* (Cambridge, MA and London, England: MIT Press, 2018).

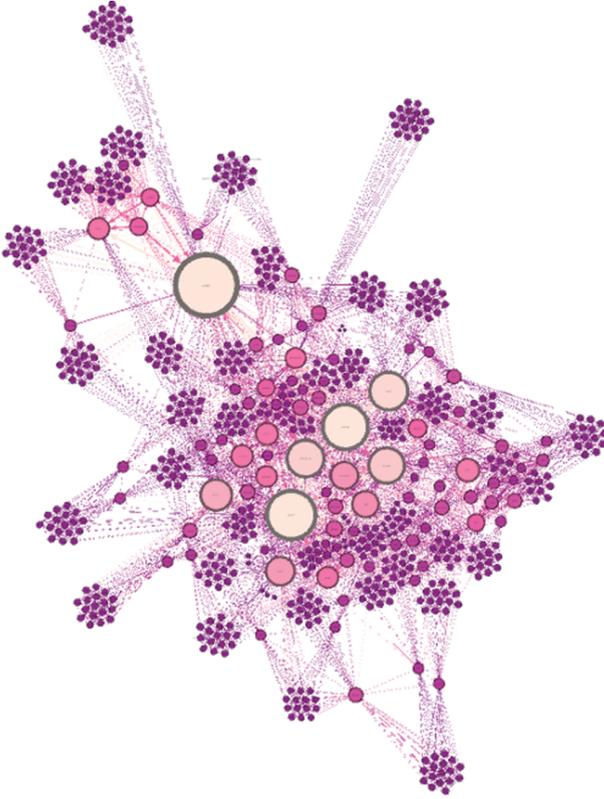


Figure 63: History of Science and Technology in the United States; color – lighter color means higher number of connections (edges) and darker color means less connections; node size – network centrality.

keyboards, etc.³⁶ Bots are one of the most important functional aspects of Wikipedia.³⁷ As shown in the networks above, they constitute the most central agents within a category-network and have connections to multiple pages. Their profiles do not include any information about their interests or education, so they cannot be easily classified, but without them the examined categories-networks would not even exist. They are not only central within one category-network but appear in all three categories, which are completely different thematically and randomly

³⁶ Tkacz, *Wikipedia and the Politics of Openness*, 111.

³⁷ Ibid. For bots on Wikipedia, also see Randall M. Livingstone, “Population automation: An interview with Wikipedia bot pioneer Ram-Man,” *First Monday* 21, no. 1 (2016), accessed July 20, 2021, <https://doi.org/10.5210/fm.v21i1.6027>

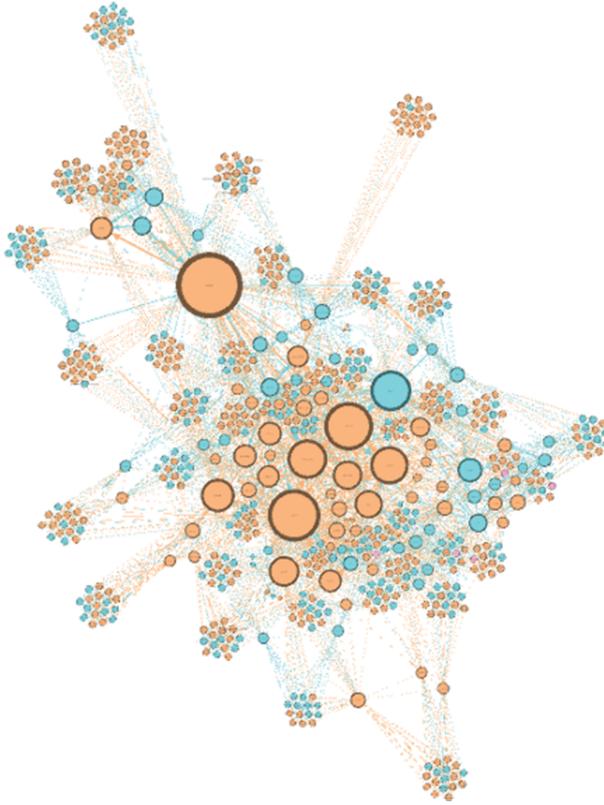


Figure 64: History of Science and Technology in the United States; in this and the following figure: color – light blue means interested in the topic they edit, orange color means no available information, and pink means not interested in the topic; node size – network centrality.

chosen. This means that these non-human actors not only have the most connections in a network, but they also determine the production of historical knowledge to a large extent. Even though they are not the editors with the highest numbers of contributions, these bots appear in all three categories. They make few edits, but these are on almost every page. While editors interested in a topic or in history make far more edits, but on fewer pages. Randall Livingstone’s argument that bots are not just programs and lines of code but represent a significant and influential population within a “sociotechnical” network of actors, is more than evident in the networks above.³⁸

³⁸ Livingstone, “Immaterial Editors,” 10.

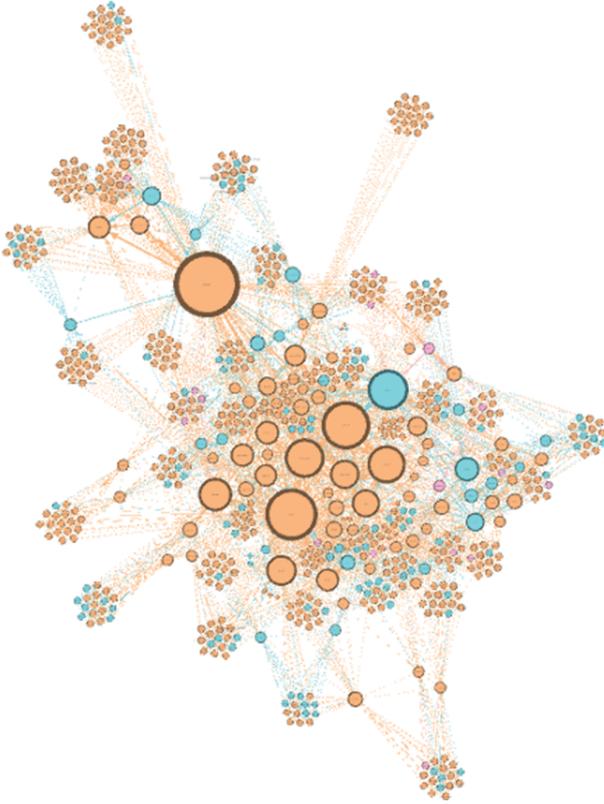


Figure 65: History of Science and Technology in the United States.

Second, bots are assigned multiple editing and administrative tasks that are essential for the production of historical knowledge. Bruno Latour was one of the first scholars in the humanities who argued that in a network there are human and non-human subjects (machines).³⁹ For him, the challenge is to investigate how these agents relate to each other and what their role is within a network. In 2008, when Wikipedia was becoming bigger, it also aimed to be more consistent in terms of checking spelling, grammar, and punctuation across all its articles. This was the major reason why Wikipedia started to introduce bots.⁴⁰ In all three categories, bots have significant administrative tasks. The “InternetArchiveBot” is one of the most central non-human agents within the three networks. Its role is to identify

³⁹ Latour, *Reassembling the Social*.

⁴⁰ Lih, *The Wikipedia Revolution*, 99.

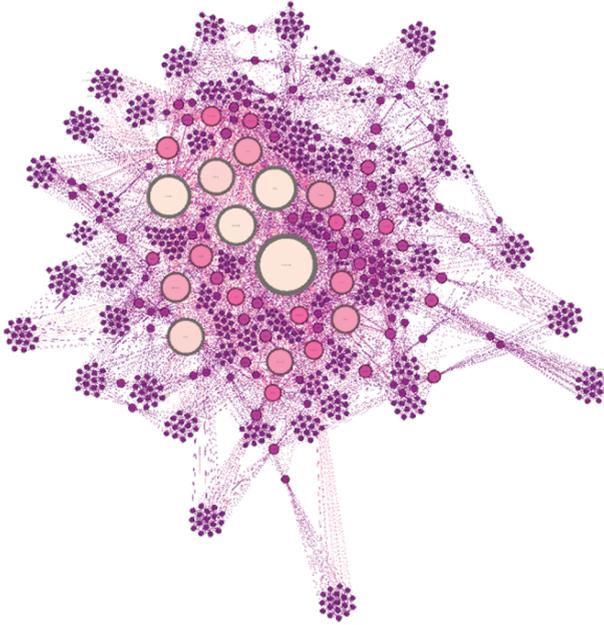


Figure 66: LGBT history in the United States; color – lighter color means higher number of connections (edges) and darker color means less connections; node size – network centrality.

and replace broken external links and make improvements to references on Wikipedia.⁴¹ It was developed by the user “Cyberpower678” and is now funded by the Internet Archive.⁴² The “ClueBot NG” is central in one of the three networks and is an anti-vandal bot, which detects and reverts vandalism quickly and automatically.⁴³ It was created and is maintained by a team of Wikipedia users.⁴⁴ “Cydebot,” also central in two of the three categories, is the “Cydebot” which is tasked with “moving and deleting categories and updating listed pages of categories”.⁴⁵ “Mon-

41 “User:InternetArchiveBot,” *Wikipedia*, accessed September 10, 2021, <https://en.wikipedia.org/wiki/User:InternetArchiveBot>; “InternetArchiveBot,” *Wikimedia*, accessed September 10, 2021, <https://meta.wikimedia.org/wiki/InternetArchiveBot>

42 “User:InternetArchiveBot”.

43 “User:ClueBot NG,” *Wikipedia*, accessed September 10, 2021, https://en.wikipedia.org/wiki/User:ClueBot_NG

44 *Ibid.*

45 “User:Cydebot,” *Wikipedia*, accessed September 10, 2021, <https://en.wikipedia.org/wiki/User:Cydebot>

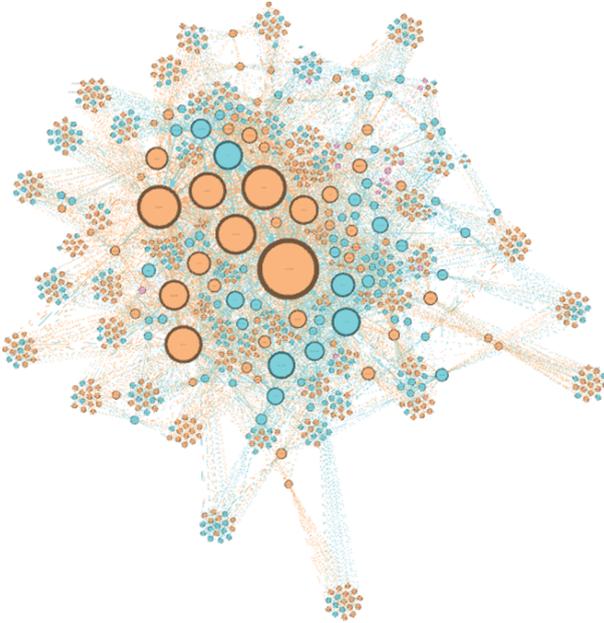


Figure 67: LGBT history in the United States; in this and the following figures: color – light blue means interested in the topic they edit, orange color means no available information, and pink means not interested in the topic; node size – network centrality.

kbot,” central in two of the three categories, makes coding edits that are tedious to do manually and is operated by the user “Trappist the monk”.⁴⁶

Another significant bot is the “AnomieBOT,” which is devoted to various tasks such as “removing pages from categories where the pages do not meet the page inclusion criteria,” “removing flag icons from infoboxes and layout templates per community consensus,” updating crosslinks when content is archived, etc.⁴⁷ It uses multiple accounts, such as “AnomieBOTII,” “AnomieBOT III,” etc., and is operated by the user “Anomie”.⁴⁸ It is central in two of the three examined networks. “KolbertBot” is tasked with the conversion of HTTP external links to HTTPS for compatible websites.⁴⁹ This conversion helps protect data integrity and

⁴⁶ “User:Monkbot,” *Wikipedia*, accessed September 10, 2021, <https://en.wikipedia.org/wiki/User:Monkbot>

⁴⁷ “User:AnomieBOT,” *Wikipedia*, accessed September 10, 2021, <https://en.wikipedia.org/wiki/User:AnomieBOT>

⁴⁸ *Ibid.*

⁴⁹ “User:KolbertBot,” *Wikipedia*, accessed September 10, 2021, <https://simple.wikipedia.org/wiki/User:KolbertBot>



Figure 68: LGBT history in the United States.

user privacy.⁵⁰ “KolbertBot” is central in one of the three networks and is run by “Jon Kolbert”. “GreenC bot” is central in two of the three categories and is responsible for multiple technical tasks, such as fixing citations, creating reports, adding maintenance tags to pages on-demand, etc.⁵¹ It is operated by the user “GreenC”.⁵² Another important bot is the “Citation bot,” also central in two networks, which is tasked with checking Wikipedia’s articles’ references, adding digital object identifiers (DOIs) to references, adding other identifiers (PMIDs, ISBNs), linking to open access repositories, and fixing formatting errors.⁵³ It is operated by “Smith609”.⁵⁴ The “RjwilmsiBot,” central in one category, has various tasks, such as tagging re-directs, completing fields of news citations, creating re-directs, correcting parame-

⁵⁰ Ibid.

⁵¹ “User:GreenC bot,” *Wikipedia*, accessed September 10, 2021, https://en.wikipedia.org/wiki/User:GreenC_bot

⁵² Ibid.

⁵³ “User:Citation bot,” *Wikipedia*, accessed September 10, 2021, https://en.wikipedia.org/wiki/User:Citation_bot

⁵⁴ Ibid.

ter names in citation templates, etc.⁵⁵ It is run by the user “Rjwilmsi”.⁵⁶ Another central bot in two categories is the “SmackBot,” which is operated by “Rich Farmbrough” and has various tasks, such as correcting grammar and spelling mistakes, replacing birth dates and death dates in infoboxes, replacing start dates and end dates, etc.⁵⁷ There are two more bots, “Bender the Bot” and “Yobot”. The former is responsible for the conversion of existing external links on Wikipedia, from an unencrypted to an encrypted transport protocol for reasons of privacy, integrity, and authentication.⁵⁸ It is operated by “bender235” and is central in two of the three case studies.⁵⁹ The latter is central in one network and is operated by the user “Magioladitis”.⁶⁰ It has multiple purposes but it is mainly focused on the categorization of individuals in categories regarding the year, date, and place of their birth/date, the time period in which they lived, their profession, etc.⁶¹ Those are just the most central bots of the three categories-networks. There are also other bots that do not have as many connections in a network but are still important for the development of Wikipedia pages.

As the roles of these bots reveal, on Wikipedia there are users who are interested either in the specific topic they have decided to create/develop or in history as a broader field. Their interest in the topic or in history encourages them to participate in the Wikipedia community, spend time researching their topic, and contribute to the production of history. Those users are mainly the editors with the highest number of contributions. However, the production of historical knowledge does not take place in a vacuum. It is not only the result of interests and passions for specific topics and fields of knowledge but also the product of a “sociotechnical” and “technomanagerial” digital environment.⁶² The major representatives of this environment are the bots of Wikipedia, along with the site’s many policies and guidelines. As José van Dijck argues, bots are “content agents” who actively engage

55 “User:RjwilmsiBot,” *Wikipedia*, accessed September 10, 2021, <https://en.wikipedia.org/wiki/User:RjwilmsiBot>

56 *Ibid.*

57 “User:Helpful Pixie Bot,” *Wikipedia*, accessed September 10, 2021, https://en.wikipedia.org/wiki/User:Helpful_Pixie_Bot This link is a redirection from the user page of “SmackBot”.

58 “User:Bender the Bot,” *Wikipedia*, accessed September 10, 2021, https://en.wikipedia.org/wiki/User:Bender_the_Bot

59 *Ibid.*

60 “User:Yobot,” *Wikipedia*, accessed September 10, 2021, <https://en.wikipedia.org/wiki/User:Yobot>

61 *Ibid.*

62 For the term “sociotechnical,” see Livingstone, “Immaterial Editors,” 10. For the term “technomanagerial,” see Niederer and van Dijck, “Wisdom of the Crowd or technicity of the content?” 1369.

with Wikipedia's contents.⁶³ The networks examined above not only confirm van Dijck's point, but also underline how central these bots are to the existence of a Wikipedia category-network. The bots have administrative roles. They do not make the same high numbers of edits that editors interested in the topic/history do on each page; however, they do make quite a few administrative contributions, and these are on almost every page of a category. While the editors interested in a specific topic/history write and develop the article further by adding content to it, bots are tasked with more technical aspects: they check grammar and spelling mistakes, revert instances of vandalism, categorize articles, upload images, check references and citations, etc. As one Wikipedia user, "Ram-Man," who operates several bots, has put it: a bot is "like the miner who produces the raw materials while some other architect/artist/builder turns it into something beautiful".⁶⁴ The architect/artists/builders are the editors, who are interested in the topic or in history and make the highest number of edits in a category-network.

However, there is another important characteristic, one strongly connected to both power and hierarchy, which is easily detectable on Wikipedia: experience. To a certain extent, the category of editors who are interested in a specific topic or history more generally is a slightly vague one. An understanding of the role of bots is clear. Bots do not have personal information, as they are non-human agents. But what about all those editors who are interested in a topic/history or those who are not? What are the main characteristics of these users? Unfortunately, most Wikipedians do not disclose adequate information about their gender, ethnicity, working position, etc. They fill their profile pages with details that they want to present to the public community of Wikipedia. Nevertheless, there tends to be a common characteristic that most editors include within their profile pages – or, even if it is not included, is usually easily detectable. This is their editing experience within the Wikipedia community. Experience is also a vague term, as it is not defined by Wikipedia. For this reason, I have chosen to define experienced editors as those users with contributions to more than ten different Wikipedia pages, or having been awarded barnstars and other editing awards by Wikipedia. In Figures 69–80, we see how many editors can be classified as experienced or inexperienced in the three categories-networks.

In Figure 69, we see that almost sixty-two percent of the editors in the category "History of organization based in the United States" are experienced and displayed in orange. Twenty-three do not have any information about their experience and are displayed in light blue. Fifteen percent are not experienced editors; namely,

⁶³ van Dijck, *The Culture of Connectivity*, 139.

⁶⁴ Livingstone, "Population automation".

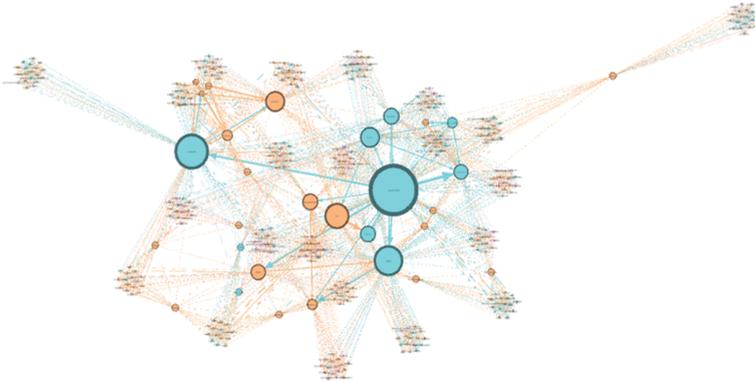


Figure 69: History of organizations based in the United States.

they have edited less than ten different Wikipedia pages and are displayed in pink within the network. As Figure 69 reveals, most users whose information about their experience is unknown tend to be bots, as they are the most central editors in the network. Some of them are also blocked, retired, or unregistered users, whose editing activities and history are not displayed on their profile pages. Regarding editors who are experienced, some of them are also central in the network, but mainly, as Figure 70 shows, the experienced editors are those with the higher numbers of edits. The network in Figure 70 does not focus on the centrality of nodes, as Figure 69 does, but on the number of contributions. The users with a significant number of edits are almost always experienced users.

Those who are inexperienced editors constitute an important number of editors in the category-network. The results are more interesting if we zoom in on the page-networks that exist within the broader category-network. In almost every page-network, there are a few inexperienced users who have only made a couple of edits (Figures 71–73). The same results appear in the other two categories. In the category “History of Science and Technology in the United States,” fifty-six percent of editors are experienced, thirty percent are marked as unknown, and fourteen percent are inexperienced (Figure 74). Most top editors are experienced and for almost every page there are a few inexperienced users, who make a few contributions (Figures 75–77). The same happens in the category “LGBT history in the United States,” where fifty percent of the involved editors are experienced, thirty-six are marked as unknown, and fourteen percent are inexperienced (Figure 78). Similarly, most top editors are experienced and a few inexperienced editors also exist for every page-network (Figures 79 and 80).

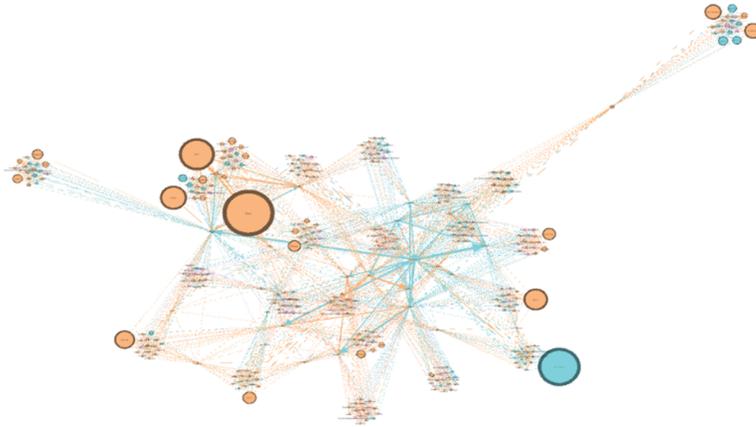


Figure 70: History of organizations based in the United States.

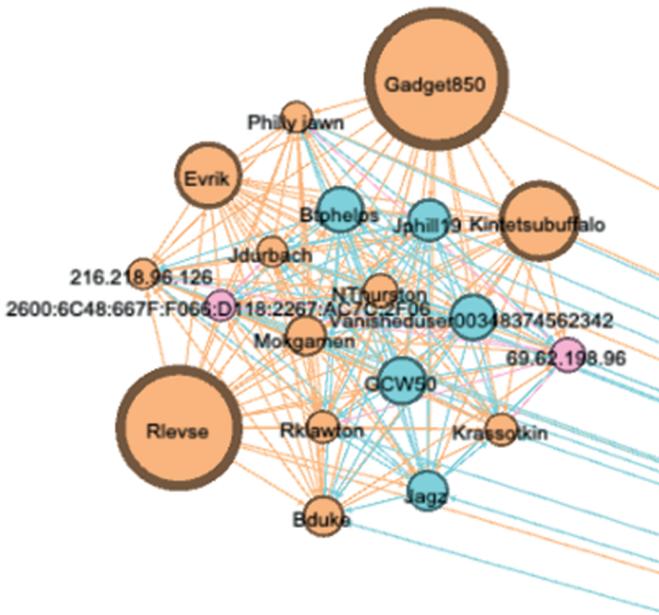


Figure 71: History of the Boy Scouts of America/History of organizations based in the United States.

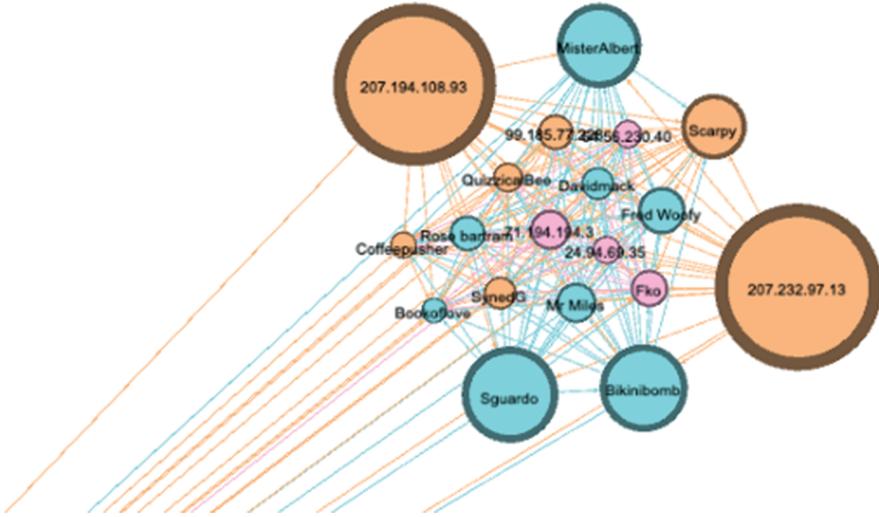


Figure 72: History of Alcoholics Anonymous/History of organizations based in the United States.

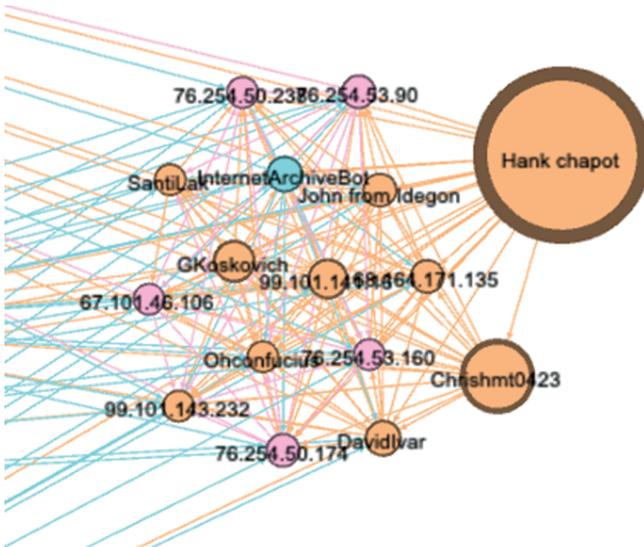


Figure 73: History of the San Francisco Police Department/History of organizations based in the United States.

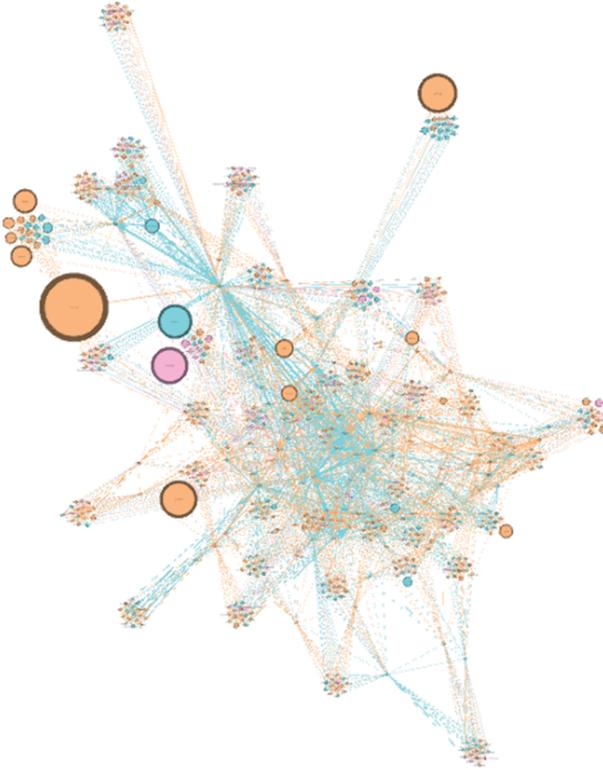


Figure 74: History of Science and Technology in the United States.

What does this repetitive pattern tell us about the editors of Wikipedia? It makes clear that there are three kinds of editors on Wikipedia: veterans, bots, and newbies. The former are experienced editors, interested either in the topic they edit or in history as a field of knowledge, and who are responsible for making the most contributions to Wikipedia. Then, we have the bots, which make only a few contributions in total but appear more frequently than any other actors in a category-network. As such, bots are the most central editors; they have the highest number of connections (edges) in a network. Their role is mostly technical and administrative but crucial for the existence of the network. Lastly, there are a bunch of editors who do not have a long history of editing experience within the Wikipedia community, but are motivated by their interest in a topic or in history itself. They make very few contributions on each page, and they are not central within the network. Typically, they only make edits in a specific article or on a topic with a broad theme, and these edits often represent their first real direct engagement with Wikipedia. These three types of editors coexist, collaborate,

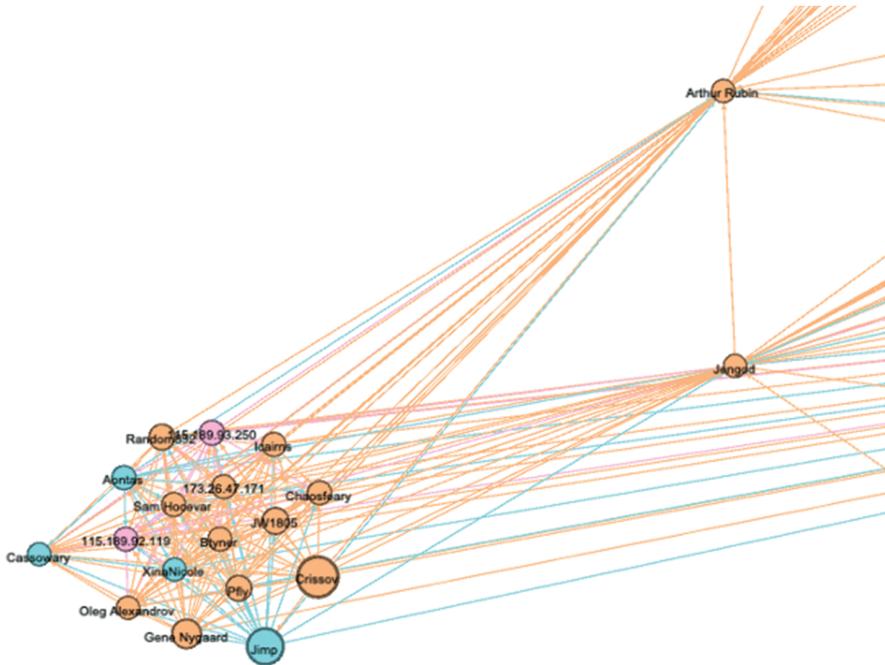


Figure 76: Plan for Establishing Uniformity in the Coinage, Weights, and Measures of the United States/History of Science and Technology in the United States.

actors appear on multiple pages within a category and even one central editor, the user “Hmains,” features across three completely different and randomly selected categories. What does this mean for the democratic nature of Wikipedia? Only a few Wikipedians make a high number of contributions to each page, while the rest of the involved editors make a significantly lower number. At the same time, a few editors are central actors in the categories-networks, as they appear more frequently in the network by editing multiple pages of a category. These two types of editors control most parts of historical knowledge production on Wikipedia.

The fact that the same patterns appear across three completely different categories-networks related to history confirms the idea that we cannot hope to fully understand Wikipedia without consideration of its technological “materiality” or “digitality”.⁶⁵ Wikipedia is not only an encyclopedic community of peo-

⁶⁵ For the term “materiality” and how it is connected to Wikipedia, see Tkacz, *Wikipedia and the Politics of Openness*, 111. For the term “digitality,” see Niels Brügger, *The Archived Web: Doing History in the Digital Age* (Cambridge, MA: MIT Press, 2018), 5.

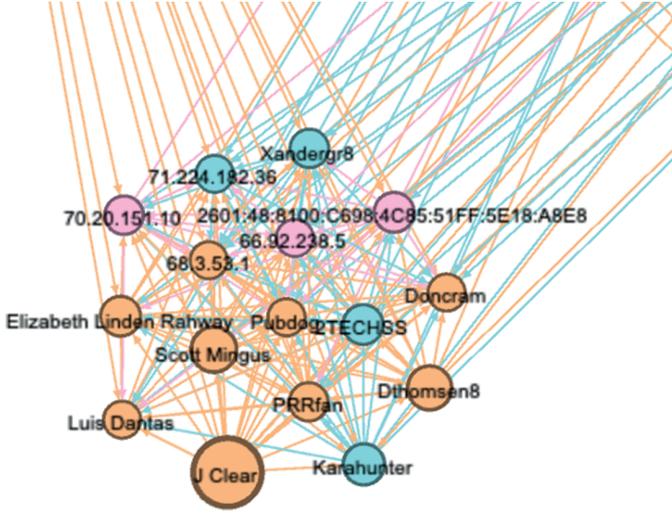


Figure 77: Phoenix Iron Works (Phoenixville, Pennsylvania)/History of Science and Technology in the United States.

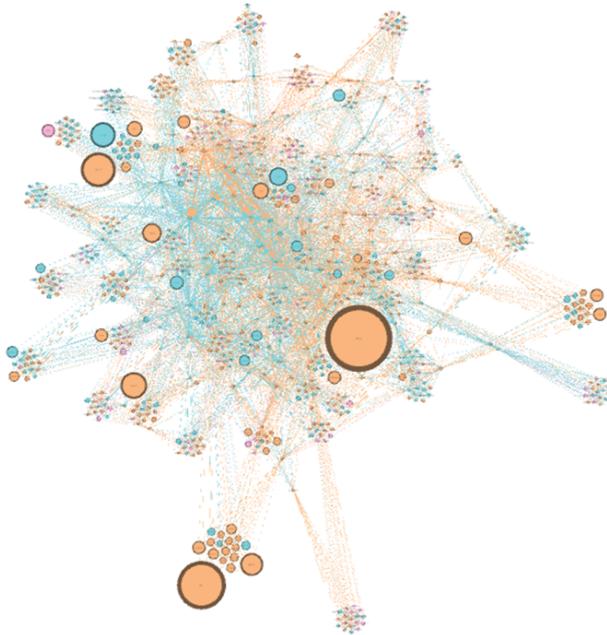


Figure 78: LGBT history in the United States.

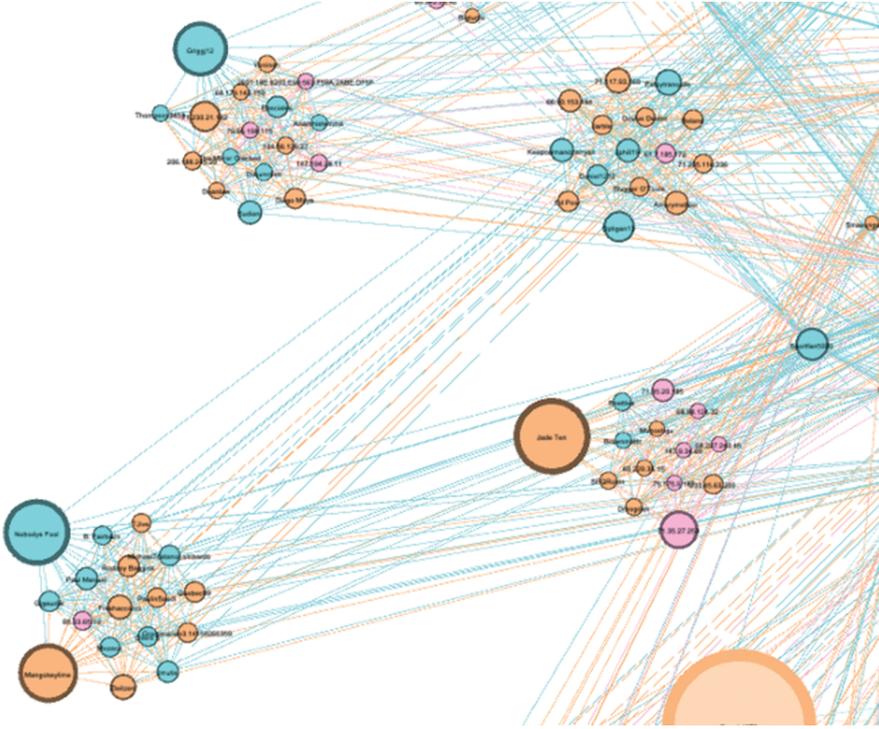


Figure 80: Multiple pages-networks/LGBT history in the United States.

experts in the specific historical areas in which they are interested.⁶⁸ On the other hand, Wikipedia is a “sociotechnical” system of power and hierarchy, which determines the production of historical knowledge. In this system, all Wikipedians, whether experienced or inexperienced, can write about history, but they do not have the same level of power to make their contributions part of the main entry.

⁶⁸ On the expertise of Wikipedians and related problems that have occurred, see Jemelniak, *Common Knowledge?*, 105–15.